Institute of Terrestrial Ecology

Centre for Ecology & Hydrology

Natural Environment Research Council

Introduction to MINITAB Release 11

by

Peter Rothery & Tim Sparks

ITE Merlewood
March 1997

Contents

1 Introduction

Minitab is a statistical package which can be used for data manipulation and a range of basic and advanced statistical techniques. Release 11 for Windows is the latest version but it also survives as Releases 7,8, 9 and 10.

This course deals mainly with the use of Minitab for data manipulation and some basic statistical methods. A brief summary of the more advanced statistical techniques and the methods new in Release 11 is also included. An associated course covers high resolution graphics.

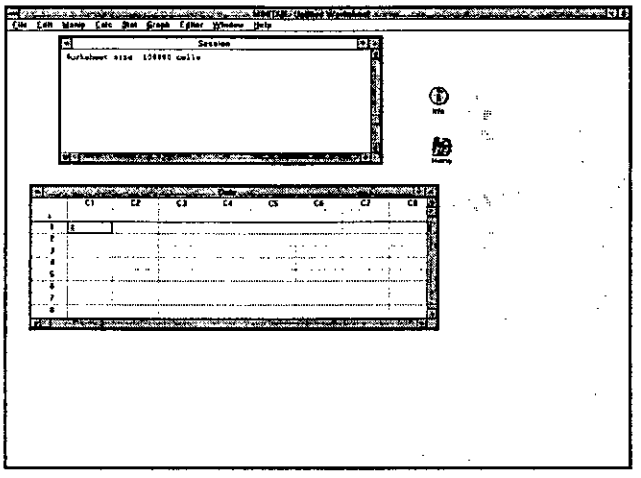
1.1 Example Dataset: Puffin Beak Measurements

For illustration we use some data on puffin beak measurements. These are lengths and depths of beaks in a sample of 41 birds (21 males and 20 females) measured in a study population on St. Kilda during 1991-93 Birds were sexed by observing behaviour because males and females are outwardly similar in size and plumage. This is very time-consuming so data were collected to see whether birds could be sexed reliably from their beak measurements. These data are held in the file puffin.dat (see Appendix).

2 Starting and Stopping Minitab

- To start: double click on the Minitab for Windows icon in the Program Manager
- To stop: choose **File - Exit** or double-click on the top left-hand button in the Program Manager

Starting Minitab gives the screen display



This consists of the following.

- A **Menu bar** to access commands for data manipulation, statistical analysis and graphics.
- A **Session window** which displays output and Minitab commands.
- A **Data window** which shows the worksheet where data can be entered, edited and viewed.
- An **Info window** (icon) which contains a summary of the worksheet
- A **History window** (icon) which contains a record of the commands used in the current session

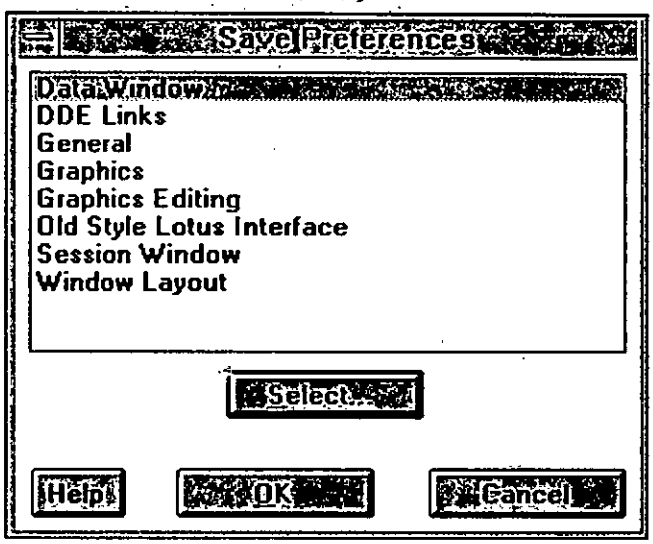
There are also **Graphics windows** for displaying high-resolution graphics.

Windows can be sized and positioned in the usual way by using the mouse. To move between windows click **Windows** from the menu bar and click the required window.

3 Customizing the Minitab Environment

Minitab can be customised to suit individual preferences as follows.

Choose **Edit - Save Preferences**
i.e. Click **Edit**, then
Click **Save Preferences**



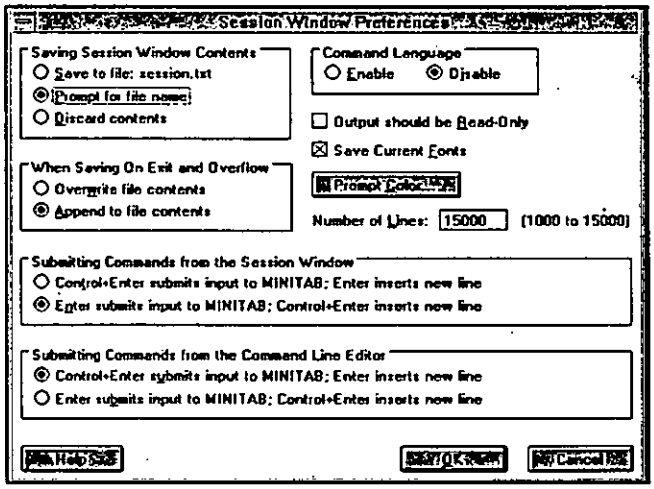
Some example settings are as follows.

3.1 Setting Session Window Options

These settings refer to the automatic saving and editing of the Session window.

From the **Save Preferences** box

1. Click **Session Window**
2. Click **Select**
3. Check the boxes as required.
3. Click **OK**, click **OK**



Prompt for file name and **Append to file contents** are the default settings and are recommend for most purposes. In Minitab 11 the Command Language is Disabled by default, for use with menus only. To use Minitab commands check **Enable**

3.2 Changing the Worksheet Size

The default size of the worksheet is 100,000 cells (i.e. data values). The maximum is 4,000,000 cells, but this may be limited by the amount of available memory on your computer. The following is guide to the maximum number of cells allowed for a given amount of RAM memory: 8Mb - 2,000,000 cells; 12Mb - 3,000,000 cells; 16Mb - 4,000,000 cells.

The current size of the worksheet is displayed in the Session window when you start Minitab. This can be changed as follows.

Choose **Edit - Save Preferences**
Click **General**
Click **Select**
Enter worksheet size in **Worksheet size at startup**
Click **OK**, click **OK**

The change will take effect the next time you start Minitab.

4 Dialog Boxes and Commands: an Example Minitab Session

Minitab 11 can be used in two ways.

- Typing commands directly into the Session window
- Accessing commands indirectly via menus and dialog boxes.

The second approach is the easier and a good way to learn Minitab. Once commands are learned, however, it is often quicker to type them directly into the Session window.

In this session we retrieve the data on puffin beak measurements into the Minitab worksheet and calculate summary statistics separately for males and females. These data have been stored as a Minitab worksheet, puffin.mtw - the method for creating Minitab worksheets will be described later. We start by retrieving the worksheet as follows.

Choose **File - Open Worksheet**

Click on the appropriate drive and directory and select puffin.mtw. Click **Proceed** if you see a message asking whether to proceed

Data

	C1	C2-T	C3	C4	C5	C6	C7
↓	Year	Sex	Length	Depth			
1	91	m	28.2000	35.5000			
2	91	f	28.5000	32.6000			
3	91	f	*	30.6000			
4	91	m	27.1000	34.3000			
5	91	m	27.8000	37.6000			
6	91	m	28.8000	39.0000			
7	91	f	28.0000	33.8000			

The worksheet shown in the Data window contains four columns of numbers for: year (C1), sex (C2), beak length (C3) and beak depth (C4), and each column is identified by a name. The asterisk for the beak length of the third bird means that the measurement is missing.

Session command

Alternatively, we could have typed the following command in the Session window after the Minitab prompt (MTB>).

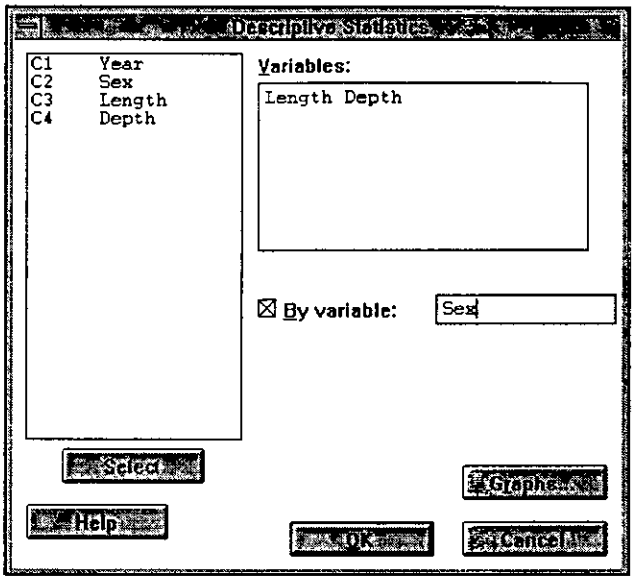
```
MTB> Retrieve 'c:\intmtb11\puffin.mtw'
```

4.1 Using the Dialog Box for Descriptive Statistics

Choose **Stat - Basic Statistics - Descriptive Statistics**

This produces a dialog box which contains a list of the columns (and names) held in the worksheet. To obtain descriptive statistics separately for males and females we fill in the dialog box as follows.

Select Length and Depth:
- click on Length and drag mouse to include Depth in the variable list
- click **Select**



Check the **By variable** box
Select Sex:
- click in the box to make it active
- click Sex in the variable list
- click **Select** button

Click **OK**

Session commands

```
MTB > Describe 'Length' 'Depth';
SUBC>  By 'Sex'.
```

or

```
MTB > Describe C3 C4;
SUBC> By C2.
```

Output

Descriptive Statistics

Variable	Sex	N	N*	Mean	Median	TrMean
Length	m	21	0	28.643	28.800	28.653
	f	19	1	28.179	28.000	28.141
Depth	m	21	0	35.343	34.800	35.247
	f	20	0	32.165	32.550	32.222

Variable	Sex	StDev	SEMean	Min	Max	Q1	Q3
Length	m	1.001	0.218	26.600	30.500	27.900	29.350

	f	0.937	0.215	26.600	30.400	27.500	29.000
Depth	m	1.417	0.309	33.500	39.000	34.300	36.300
	f	1.686	0.377	28.600	34.700	30.875	33.650

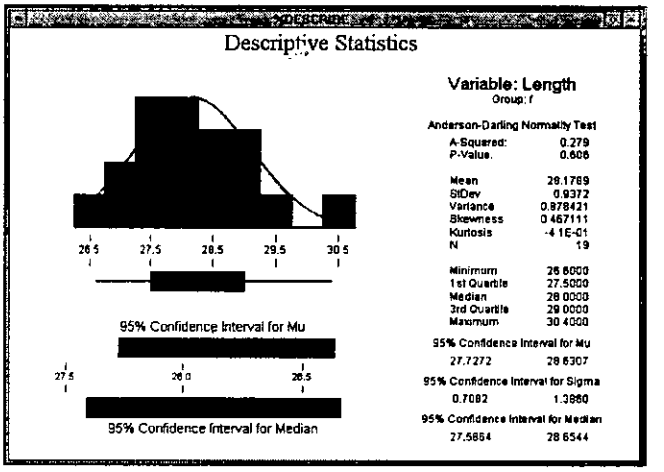
Interpretation

The statistics are as follows: Total number of non-missing values (N); Number of missing values (N*); Arithmetic mean (Mean); Trimmed mean (Trmean) - calculated by removing the smallest and largest 5% of values; Standard deviation (StDev); Standard error of the mean (SEMean); Minimum (Min); Maximum (Max); Lower quartile (Q1); Upper quartile (Q3). .

The results suggest that males have deeper beaks than females but that the lengths are similar.

Graphical output

In the the **Descriptive Statistics** dialog box click on **Graphics** for a list of graphical outputs including Histogram, Histogram+Normal Curve, Dotplot, Boxplot and Graphical Summary which gives summary statistics together with the Histogram+Normal Curve as shown for beak lengths of females.



Structure of Session Commands

This Session command consists of the command Describe completed by the semi-colon (;) and the subcommand By. More generally, Minitab commands can have several subcommands. The complete sequence of command plus subcommands is ended by a period (.). Session command names are not case-sensitive and they can be abbreviated to the first four letters (except in %Macros, see section 13). Commands can be continued on the next line by ending the line with the & symbol (this produces the prompt CONT>).

4.2 Edit Last Command Dialog Feature

Choose **Edit - Edit Last Command Dialog**

This recalls the last dialog box used in the current session. It can save time when minor changes are needed, when developing code or experimenting with graphics.

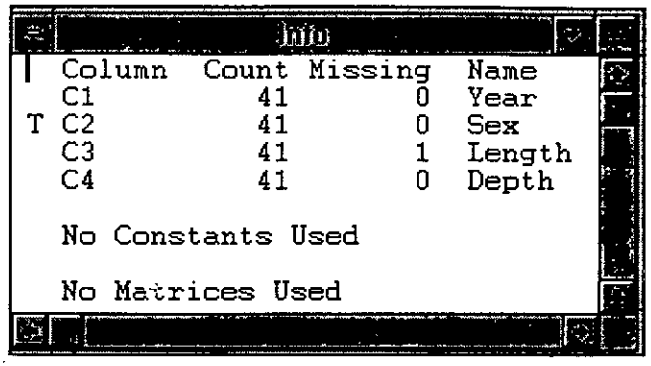
4.3 Resetting Dialog Box Defaults

Minitab keeps the settings of all dialog boxes the same as the last time that they were used in the current session. To reset the dialog box to its default state press the function key **F3**.

4.4 Info Window

To see a summary of the current worksheet:

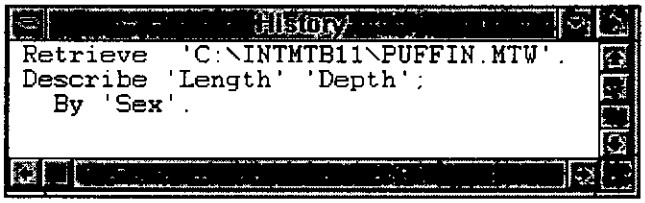
Choose **Window** and click on **Info**, or double-click on the **Info** icon



4.5 History Window

The History window gives a record of the commands used in the current session, but not including the output. To view this

Choose **Window** and click on **History**, or double-click on the **History** icon



4.6 Clearing the Worksheet and Restarting Minitab

Choose **File - New Worksheet** to clear the current worksheet.

Choose **File - Restart Minitab** to start a fresh session without leaving Minitab.

5 Help

To get help in Minitab

- Choose a command in the **Help** menu
- From any dialog box, click the Help button in the lower left hand corner.

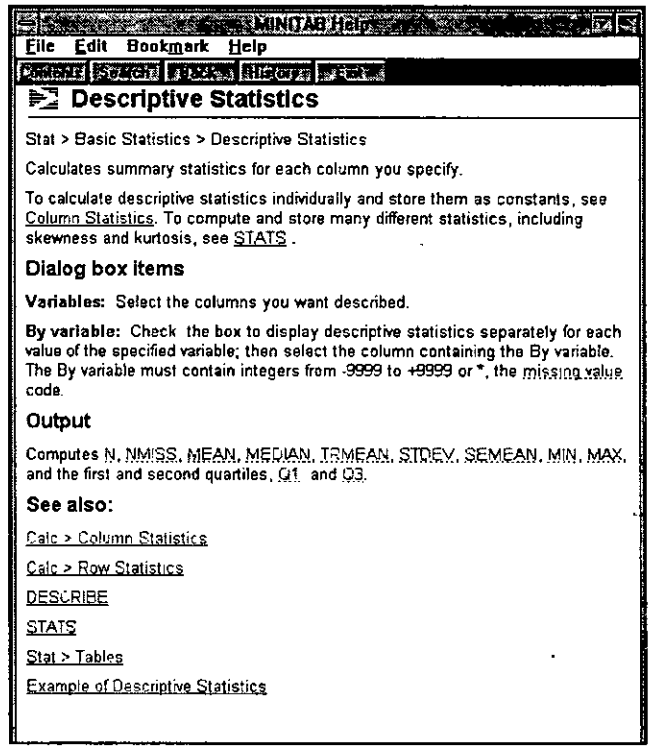
For example, to find out more about the DESCRIBE command

Choose **Stat - Basic Statistics - Descriptive Statistics**
Click the Help button

Choose from the menu to:
- Print a Help topic (File menu)
- Copy text from the Help screen to the Clipboard (Edit menu)
- Add notes to a Help topic (Edit menu)
- Keep the Help window visible at all times (Help menu)

Click on the buttons to:
- Go to a table of Contents
- Search for a topic in the index
- Goto to the Help screen last viewed (Back)
- Goto any Help screen viewed in the current session (History)

Click green underlined text to go to a different Help screen or to view a topic.



To exit Help, choose **File - Exit**, or double-click the box in the top left corner of the Help screen.

Contents of Help can be cut and pasted to the Notepad or a Word Processor such as WordPerfect or Word.

6 Getting Data into Minitab

This section deals in more detail with the structure of the Minitab worksheet, different ways of getting data into it, and editing the worksheet.

The main ways of entering data into the worksheet are (a) type it; (b) read it from a file; (c) paste it from another application; (d) generate it.

File - Open Worksheet is a useful general method for getting data into the worksheet. Further details and some other methods are described below. First, we discuss the different types of data and the structure of the worksheet

6.1 Types of Data

Mintab can handle three types of data which can be entered, imported, exported, converted to other types and used in data manipulations and calculations.

Numeric data

These are just numbers. They can be stored in decimal point format (e.g 123.45) or in exponential format (e.g. 1.2345e+02). The largest (smallest) number is +/-1e+18), and when this is exceeded Minitab converts the number to a missing value (*) and returns a message.

Text data

Text data can consist of a mixture of letters, numbers, spaces and special characters. For example, in the puffin data males and females are coded as m and f. When text data is entered into a column in the worksheet a letter T appears after the column name (e.g. C1-T or Sex-T).

Date/Time data

Data can be stored as dates (e.g. Jan-17-1997 or 1/17/97), as times (08:50:35 am), or as a mixture of the two (1/17/97 08:50:35 am). Columns containing data/time data have a letter D after their name. Date/time variables appear in date/time format but they are stored as numbers so that they can be used in calculations that accept numbers. The internal number consists of an integer which represenst the number of days since December 30 1899, and a fraction of the day completed. Dates and times before 30/12/1899 are stored as negative numbers.

The data/time format can be changed (e.g. from Jan-17-1997 to 17-Jan-1997 or 17/1/97) as follows.

1. Select one or more columns which must be empty or already in date/time format
2. Choose **Editor - Format Column - Date/Time**
3. Select a format from **Current Date/Time Formats** box.

Date/time variables can be used to subset data and as factors or classification variables in cross tabulation, calculation of descriptive statistics and analysis of variance. For use of Date/Time data on graphs see Introduction to Graphics in Minitab.

6.2 Worksheet Structure

Columns

The worksheet for the puffin beak measurement data consists of four columns: Year (C1); Sex (C2); Length (C3); Depth (C4). This way of storing a variable, such as length, in one column and a classification factor, such as sex, in another is particularly convenient for many of the analyses which can be carried out using Minitab. However, different columns could be used to store data for different groups of individuals - for example, data for males and females could be stored in different columns. A Minitab worksheet can contain up to 1000 columns.

Constants

The worksheet can contain up to 1000 stored constants (denoted by K1, K2,..,K1000, or by name). Stored constants do not appear in the Data window but they are listed in the Info window and can be printed in the Session window. Minitab assigns the following: K998 = * (missing value); K999 = 2.71828 (*e*); K1000 = 3.14159 (pi), but these can be changed.

Matrices

Minitab can store up to 100 matrices (denoted by M1, M2,...,M100, or by name). These are not shown in the Data window but appear in the Info window.

Naming columns, constants and matrices

```
MTB > Name c10 'name'
```

This assigns a name to a column (constant or matrix) which can be up to to 31 characters long.

Size of Worksheet

When Minitab is started the size of the worksheet is displayed in the Session window. The default size of the worksheet is 100,000 cells (i.e. data values) and the maximum is 4,000,000 cells, but this may be limited by the amount of available memory on your computer. To change the size of the worksheet use **File - Save Preferences** (see section 3.2).

6.3 Entering and Editing Data Directly in the Data Window

Entering and editing data in the Data window is similar to other spreadsheet applications. A summary is as follows.

- To open Data Window: choose **Window - Data** or click in Data Window
- To make cell active: click on cell (active cells have a dark surround)
- To enter data: type value in active cell and press Enter
  (The Data Entry Direction Arrow in the upper left-hand corner points to the direction moved after pressing the Enter key. Click on it to change the direction)
- To return to the beginning of the next row (column): press Crtl+Enter when the direction arrow points right (downward)
- To select row (column): click on row (column) number
- To select a rectangular area: click on first cell and drag to last cell
- To delete row (column, area): select and press Del
- To insert row/column: choose **Editor - Insert Rows/Columns**
  (the row/column is inserted before the highlighted row/column)
- To copy cells to clipboard: select cells and choose **Edit - Copy Cells**
  (or use CTRL+C to copy cells)
- To paste cells from clipboard: choose **Edit - Paste/Insert Cells**
  (or use CTRL+V to paste cells)

***** It is wise to save the worksheet frequently *****

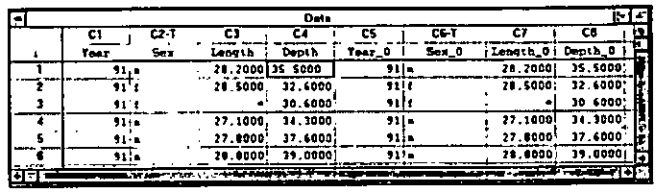
Choose **File - Save Worksheet** or **File - Save Worksheet As.** (see Section 7 on saving, exporting and printing Data)

****

6.4 Retrieving a Minitab Worksheet

Choose **File - Open Worksheet**

Select drive, directory and file
(this overwrites the existing worksheet)



Data

	C1	C2-T	C3	C4	C5	C6-T	C7	C8
	Year	Sex	Length	Depth	Year_0	Sex_0	Length_0	Depth_0
1	91	m	28.2000	35.5000	91	m	28.2000	35.5000
2	91	f	28.5000	32.6000	91	f	28.5000	32.6000
3	91	f	*	30.6000	91	f	*	30.6000
4	91	m	27.1000	34.3000	91	m	27.1000	34.3000
5	91	m	27.8000	37.6000	91	m	27.8000	37.6000
6	91	m	28.8000	39.0000	91	m	28.8000	39.0000

or

Choose **File - Merge Worksheet**

(this adds data to the existing worksheet).

If some of the columns are common to both worksheets then Minitab will rename them by adding a index number. For example, if the puffin worksheet puffin.mtw is merged with itself the resulting worksheet is as shown with renamed columns Year_0 Sex_0 etc.

6.5 Using Copy and Paste to Import Data from Other Applications

Data can be copied and pasted from other applications (e.g. the spreadsheet Excel) as follows.

1. In the application , highlight the data and copy to the clipboard
   (Select data and use **Edit - Copy** or CTRL+C)

2. In Minitab, copy the new data into the worksheet
   (Select data and use **Edit - Paste/Insert Cells** or CTRL+V) .
   Minitab places the data at the beginning of the column containing the active cell.

6.6 Dynamic Data Exchange (DDE) - Minitab as a Client

Dynamic Data Exchange can be used to transfer data between Minitab and other applications that support DDE. Minitab can act as a *client* (receiving data), or as a *server* (sending data). With Minitab as a client (receiving data) a link is established in which changes made in the server application are automatically updated in the Minitab worksheet (but not vice versa). The default can be changed using **Edit - Save Preferences - DDE Links** and checking **Ignore external data updates**. To import data and set up the link follow steps 1 and 2 in section 6.5 using **Edit - Paste Link** to copy the data into the Minitab worksheet.

6.7 Importing Data from an ASCII File

There are two ways to import data from an ASCII file (DAT, TXT)
File - Open Worksheet and **File - Other Files - Import Special Text.**

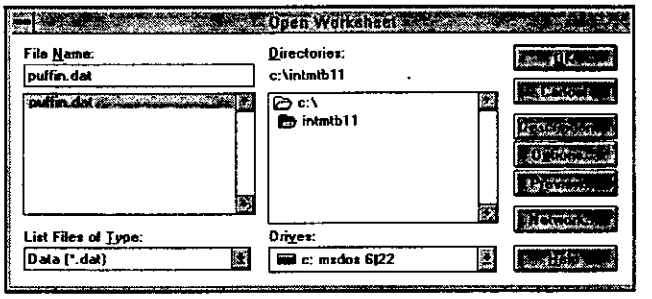
For illustration we use the file puffin.dat which contains four columns of numbers for year, sex, beak length and beak depth, each number separated by one or more spaces.

Method 1

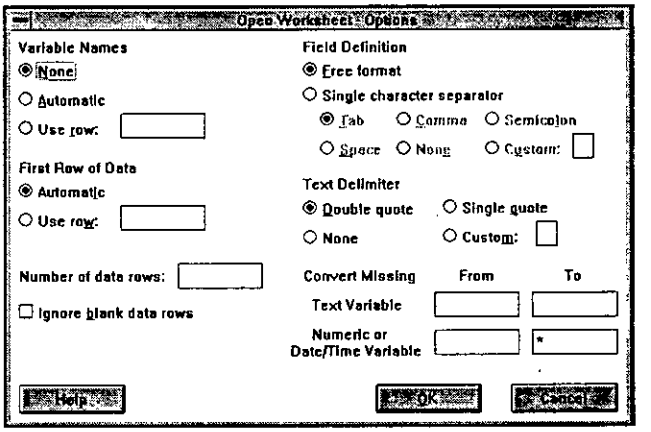
Choose **File - Open Worksheet**

Select the drive, directory and data file.
Click on **Options** to move to the

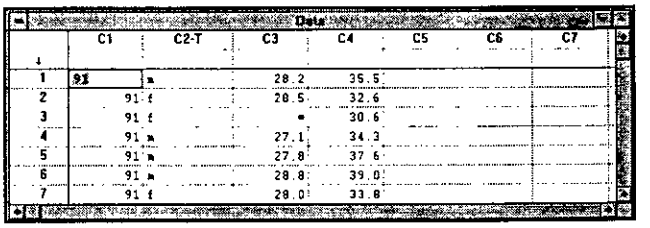
Open Worksheet - Options box



Click **None** in **Variables Names** box
(When **Automatic** is checked Minitab assigns the first row of data as column names).
Check **Automatic** in **First Row of Data** box to begin reading data from the first row.
Check Free format in **Field Definition** box.
Click **OK** to return to
Open Worksheet box.
Click **OK**



The contents of the worksheet are as shown.
Names can be added to the columns by editing the worksheet.



Columns of unequal length

When the columns are of unequal length Minitab will fill out the worksheet with the missing

values (*) to make all columns of equal length.

Previewing the worksheet

Choose **File - Open Worksheet - Preview**

Preview is used to inspect the contents of a file before reading it in. It can also be used to specify additional instructions for reading in the data thereby increasing flexibility to handle different types of data layout. This can be useful when the file contains text which needs to be skipped.

For example, the text file puffin.txt contains the puffin beak measurement data with some comments describing the contents of the file. The file can be previewed and comment lines skipped as follows.

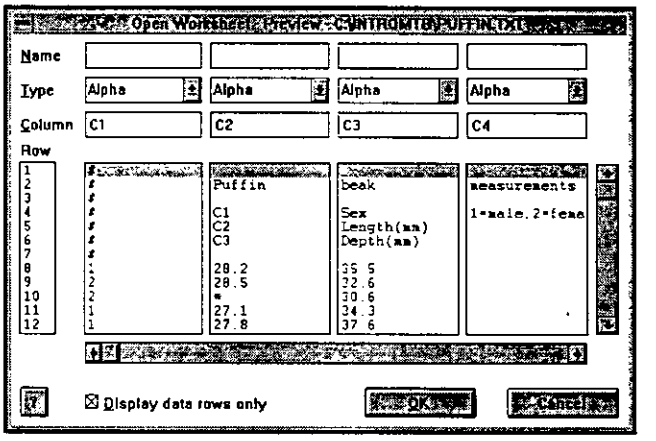
Choose **Open Worksheet.** Select disk, directory and puffin.txt.
Click on **Options**. Check **None** in **Variables Names** box.
Click on **Automatic** in **First Row of Data** box. Click **OK** (to return to **Open Worksheet**)

Click on **Preview**

This shows how the data would appear in the worksheet. The comments on file would be read into text columns. To read numeric values to named columns proceed as follows.

Change text columns to numeric in **Type** boxes where appropriate
Click **OK**

From **Open Worksheet** box:
Click **Options**
Check **Use Row** of
First Row of Data box.
Click on box and type in 8.
Click **OK**. Click **OK**.



The data are then entered into the worksheet as shown above.

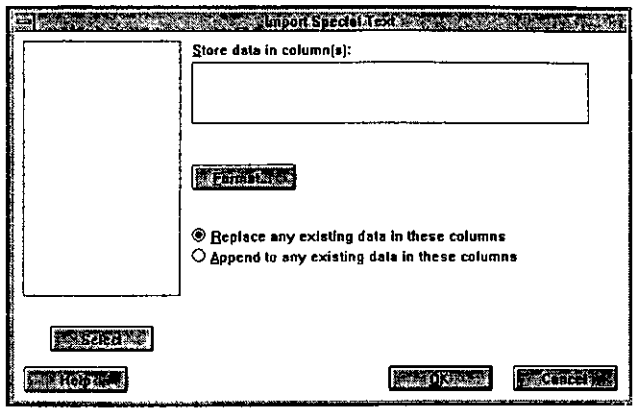
Method 2

Choose **File - Other Files - Import Special Text**

This works for numeric data only.

To read data into specified columns type column names in the
Store data in columns box
Click **OK**
(Returns to **Import Text From File** box)

Select drive, directory and file.
Click **OK**



The **Import Special Text** dialog box has a **Format** option which allows Fortran-type format statements. This could be used to read in text data in fixed format. The corresponding session command is READ which has a subcommand FORMAT.

Columns of unequal length

If the input columns are of unequal length then complete rows will be read into the worksheet but incomplete rows will be entered as missing values.

6.8 Importing Data from Spreadsheets

Choose **File - Open Worksheet**

This can be used to copy files into Minitab from a range of spreadsheets. According to the manual, Minitab 10 supports the following: Microsoft Excel (XLS); Borland Quattro Pro (WB1, WQ1); Lotus 1-2-3 (WP4, WK3, WK2, WK1, WKS); Lotus Symphony (WR1, WRK); dBase (DBF).

In some cases it might be easier to copy and paste data from the application (Section 6.4).

7 Saving Exporting and Printing Data

Minitab has several commands for saving and exporting the contents of the Session, Data and History windows. A useful general method is

Choose **File - Save Window As**

This saves the contents of the active window to a specified file.

Copy and **Paste** can also be used to export the contents of an active window and to move data and commands between windows within in Minitab.

More details of saving, exporting and moving data are as follows.

7.1 Saving the Worksheet

Choose **File - Save Worksheet** or choose **File - Save Worksheet As**

Select file name (e.g. c:\intmtb11\puffin.mtw)

Session command

```
MTB > Save 'c:\intmtb11\puffin.mtw';
SUBC> Replace.
```

This saves the worksheet in a file. Minitab worksheets are stored as files with extension MTW, or as portable worksheets with extension MTP. Release 11 worksheets should be stored in portable format for use with Minitab Release 8 or earlier.

Save Worksheet As can be used to store the worksheet different of formats. According to the manual, Minitab 10 supports the following: Microsoft Excel (XLS); Borland Quattro Pro (WB1, WQ1); Lotus 1-2-3 (WP4, WK3, WK2, WK1, WKS); Lotus Symphony (WR1, WRK); dBase (DBF).

Save Worksheet As can also be used to save data in a text file as described in the section on exporting ASCII files.

7.2 Using Copy and Paste to Export Data to Other Applications

Data can be copied and pasted to other applications (e.g. the spreadsheet Excel) as follows.

1. In Minitab , highlight the data and copy to the clipboard
   (Select data and use **Edit - Copy** or CTRL+C)

2. Copy the data into the spreadsheet
   (Select data and use **Edit - Paste/Insert Cells** or CTRL+V) .

7.3 Exporting Data to an ASCII File

There are two ways to export data to an ASCII file.

Method 1

Choose **File - Save Worksheet As**

Specify the drive, directory and file name (.TXT or .DAT). This has some options not available with Method 2 - for example, variable names can be included in the file.

Method 2

Choose **File - Other Files - Export Special Text**

Select columns to export in **Columns to Export** box. Click **OK**. Select file name (.DAT or .TXT). Click **OK**. The **Export Special Text** dialog box has options to specify a Fortran format statement to allow more control over the layout.

Session command

```
MTB > Write 'c:\intmtb11\puffin.dat' 'Year' 'Sex' &
'Length' 'Depth';
SUBC> Replace.
```

Note the use of the continuation symbol (&) to continue the command onto the next line

7.4 Saving and Editing the Session Window

In some cases the simplest method of saving output from the Session window is to cut and paste it to the Notepad or into a Word Processor such as WordPerfect or Microsoft Word.

Two other methods described below also illustrate editing the Session window and controlling the printing of Minitab commands in the Session window. In each case a file is created which

can be edited using a word processor. The Minitab output is written in a non-proportional font so that all printed characters are the same width. If the word processor is using a proportional font, the output from Minitab will not be lined up correctly. To remedy this, change the appropriate section of the document to a non-proportional font such as Courier.

Method 1

Choose **File - Save Window As**

When the Session window is active this saves the contents of the Session window (from the start of the current session) in a text file.

By default the session window can hold up to 15,000 lines at one time but this can be changed by choosing the **Session Window** option of **Edit - Save Preferences**. Or, to avoid losing output, **Edit - Save Preferences** can be used to tell Minitab to prompt for a file name in which to store the session.

Editing the session window

Before saving the contents of the Session window they can be edited as follows.

Choose **Editor - Make Output Editable**. If the Editor menu says **Make Output Read-Only**, there is no need to do anything.

To set the Session window output to Read-Only, choose **Editor - Make Output Read-Only**.

Method 2

At any time during a Mintab session the session commands OUTFILE and NOOUTFILE can be used to open and close a file to store the contents of the Session window, as follows

```
MTB > Outfile 'c:\intmtb11\puffin.lis'
  .
  .
MTB > Nooutfile
```

Disabling and Enabling the Command Language

Minitab commands generated by dialog boxes can be suppressed so that they do not appear in the Session window. To do this choose **Editor - Disable Command Language**. To restore the command prompt and reflect subsequent commands, choose **Editor - Enable Command Language**. When commands are disabled there is no prompt and commands cannot be typed into the Session window. In Minitab 11 the commands are disabled by default but this can be

changed using **Edit -Save Preferences - Session Window** (section 3.1).

7.5 Copying Session Window Output to the Worksheet

Sometimes the output in the Session window may be required for further analysis, e.g. means produced by the DESCRIBE command. Output can be copied as follows.

1. Highlight the required data in the Session window by holding down both mouse buttons and dragging.

2. Copy the data using **Edit - Copy** or Ctrl+C.

3. In the Data window, locate the cursor in the top left-hand corner of the target area for the data and paste using **Edit - Paste** or Ctrl+V.

When pasting several columns side-by-side a dialog box appears with options on spacing.

7.6 Saving Minitab Commands

Keeping a copy of Minitab commands can be useful for rerunning analyses and for developing macros (see section on macros). Commands can be cut and pasted to the Notepad or from the Session and History window to the Session window. They can also be saved on file using the methods described below. Method 1 is quicker but Method 2 allows more flexibility.

Method 1

Choose **File - Save Window As**

When the History window is active this saves the commands, from the start of the current session, in a specified text file.

Method 2

This uses the session commands JOURNAL and NOJOURNAL to open and close a file to save the commands at any time during a Minitab session, as follows

```
MTB > Journal 'c:\intmtb11\puffin.mtj'
.
.
MTB > Nojournal
```

8 Data Manipulation

The **Manip** menu contains commands for copying and subsetting, stacking and unstacking, coding data and changing data type, ranking and sorting columns of data, and printing data stored in columns, constants or matrices. It can also be used for deleting rows of in a column and for erasing columns, although in some cases it might be easier to edit the worksheet directly.

To illustrate the commands, it is convenient to use a small set of data so that the results of the manipulations can be easily viewed. We use a worksheet consisting of the first 7 lines of the puffin data.This could be created by using the editorto delete rows 8-41 of puffin.mtw
or by using the DELETE command as described below.

8.1 Delete Rows

Choose **Manip - Delete Rows**

Fill in the dialog box to delete the required rows.

Data

	C1	C2-T	C3	C4	C5
↓	Year	Sex	Length	Depth	
1	91	m	28.2000	35.5000	
2	91	f	28.5000	32.6000	
3	91	f	*	30.6000	
4	91	m	27.1000	34.3000	
5	91	m	27.8000	37.6000	
6	91	m	28.8000	39.0000	
7	91	f	28.0000	33.8000	
8					

Session command

```
MTB> Delete 8:41 'Year' 'Sex' 'Length' 'Depth'
```

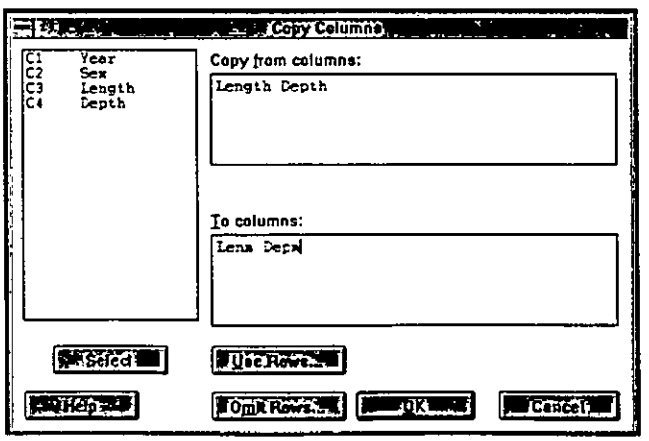
8.2 Copying and Subsetting Data

Choose **Manip - Copy Columns**

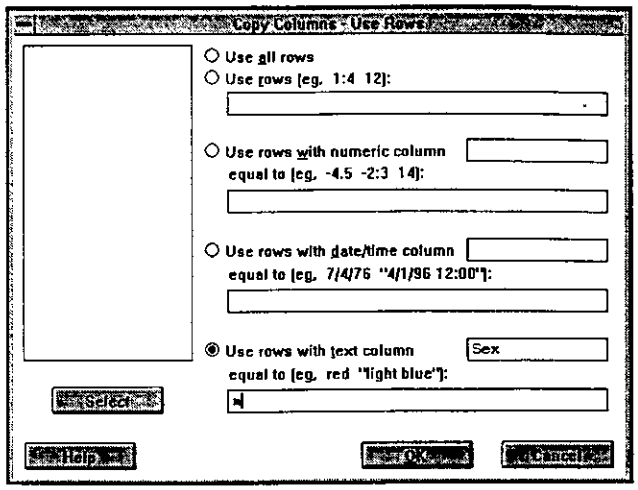
This is used to select subsets of data based on the values held in a specified column which can be numeric, date/time or text. Some examples are as follows. A copy of the resulting worksheet is given only for the first example to save space.

Copy data for males only

Select Length and Depth to copy
Click **To columns** (make active)
Type in Lenm Depm
(names for columns)
Click **Use Rows**
Moves to **Copy - Use Rows** dialog box



Check **Use rows with text column**
Click small box (make active)
Select Sex
Click large box (make active)
Type in m
Click **OK**



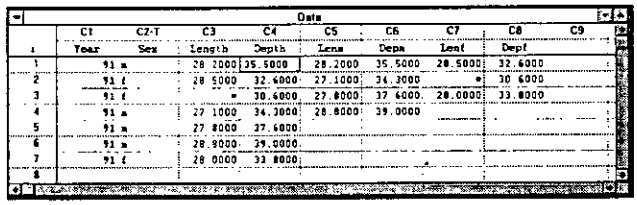
Returns to Copy Columns dialog box

Click **OK**

Session commands

```
MTB> Name C5 'Lenm' C6 'Depm'
MTB> Copy 'Length'  'Depth'  'Lenm'  'Depm'
SUBC> Use  'Sex' = m.
```

Similarly, data for females can be copied into columns Lenf and Depf to produce the worksheet as shown.



Further examples of Copy - Use Rows and Copy - Omit Rows

1. Copy data for particular rows

Check **Use rows with numeric column**. Click box to make active. Type in row numbers (e.g. 1 3 4, or 1:4 as shorthand for 1 2 3 4 etc). Click **OK**, click **OK**.

2. Copy data for beak depths between 27-28mm

Check **Use rows with column** . Click box and select Depth. Click large box and type in 27:28. Click **OK**, click **OK**

To copy data omitting particular rows, choose **Copy - Omit Rows** from **Copy** and proceed as above.

3. Copy data omitting rows with values of Length missing

Check **Omit rows with numeric column.** Click box and select Length. Click large box and type in *. Click **OK**, click **OK**.

*** **Important** ***

When using **Copy - Use** and **Copy - Omit** in the same session, check to make sure that the settings in both boxes are as required. Previous settings can lead to unexpected results. For example, when using **Copy - Omit**, make sure that the **Use all rows** of **Copy - Use** is checked to avoid additional restrictions.

****

8.3 Stacking Columns

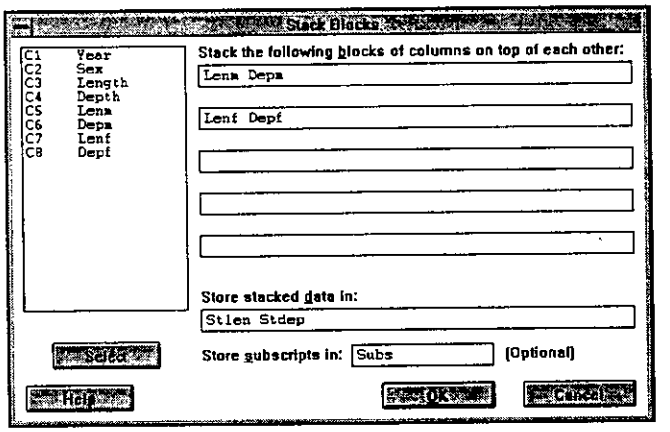
Choose **Manip - Stack/Unstack - Stack** or **Manip - Stack/Unstack - Stack Blocks**

This stacks columns or blocks of columns on top of each other.

To illustrate this start with the worksheet consisting of the first 7 lines of the puffin data with data for males in Lenm (C5) and Depm (C6) and those for females in Lenf (C7) and Depf (C8). To stack lengths and depths for males and females, and produce a corresponding column containing the sex code (subscript), fill in the dialog box as follows.

Select Lenm and Depm for the first **Stack the following blocks** box
Select Lenf and Depf for second box
Click on **Store results** in box

Type in Stlen Stdep
Check **Store subscripts in** box
Type in Subs
Click **OK**



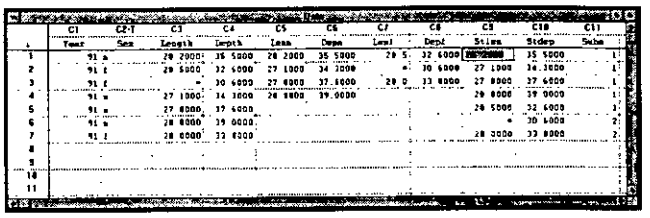
The dialog box allows for up to 5 blocks of columns. With more blocks use session commands.

Session commands

```
MTB > Name  C9 'Stlen' C10 'Stdep' C11 'Subs'
MTB > Stack ('Lenm' 'Lenf')  ('Depm' 'Depf')  ('Stlen' Stdep');
SUBC> Subscripts 'Subs'.
```

The brackets can be omitted if there is only one argument.

The worksheet is then as shown



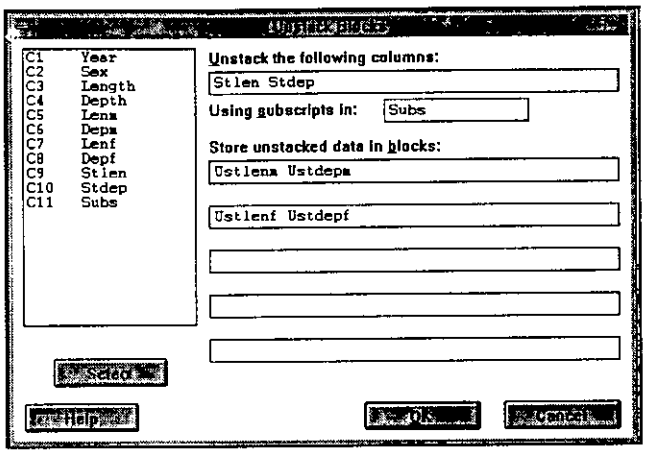
8.4 Unstacking Columns

Choose **Manip - Stack/Unstack - Unstack** or **Manip - Stack/Unstack - Unstack Blocks**

This unstacks columns or blocks of columns according to the values of a specified group variable.

The steps are essentially the reverse of those for stacking columns.

To unstack the columns in the above illustration of STACK fill in the dialog box as shown.

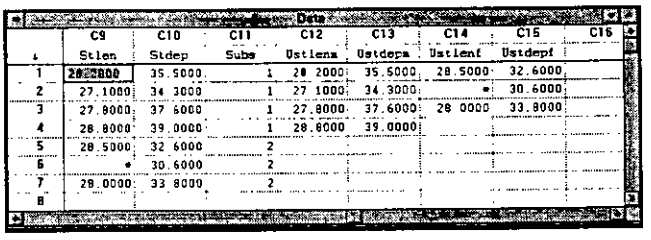


Session commands

```
MTB > Name C12 'Ustlenm' C13 'Ustdepm' C14 'Ustlenf' &
CONT>      C15 'Ustdepf'
MTB > Unstack ('Stlen' 'Stdep') ('Ustlenm' 'Ustdepm')  &
CONT>  ('Ustlenf' 'Ustdepf');
SUBC> Subscripts 'Subs'.
```

The continuation symbol & means that the command is continued on the next line.

The worksheet is then as shown.



8.5 Coding Data Values

Choose **Manip - Code**

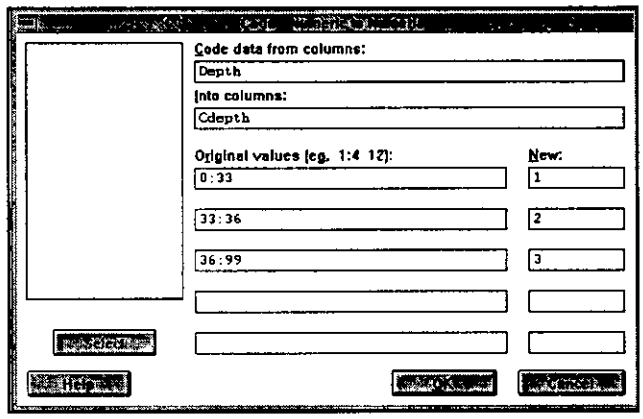
This codes or recodes values in a column according to specified values or a range of values. It allows the following combinations: Numeric to Numeric; Numeric to Text; Text to Text; Text to Numeric. There is also a Conversion Table option to define a coding scheme for Old and New values.

For example, to code puffin beak depths into following size classes.

Depth =< 33 coded as 1; 33 < Depth =< 36 coded as 2; Depth > 36 coded as 3.

Fill in the dialog box as follows.

Select Depth for
Code data from columns box.
Click on **Into columns** box.
Type in Cdepth.
Click on first.
Original values box. Type in 0:33.
Click on **New** box. Type in 1.
Click on second box. Type in 33:36.
Click on **New** box. Type in 2.
Click on third box . Type in 36:99.
Click on **New** box. Type in 3.
Click **OK**



The last group upper limit of 99 will definitely include all birds ! The colon is used to indicate values lying within a range. By convention: 0:33 is greater than zero and less than or equal to 33, whereas 33:36 is strictly greater than 33, and less than or equal to 36.

The dialog box allows only five groups. For more groups use the Session command.

Session commands

```
MTB > Name  C5 'Cdepth'
MTB> Code (0:33) 1 (33:36) 2 (36:99) 'Depth' 'Cdepth'
```

The first few lines of the worksheet are then as shown.

Data

	C1	C2-T	C3	C4	C5	C6	C7	C8
	Year	Sex	Length	Depth	Cdepth			
1	91	m	28.2000	35.5000	2			
2	91	f	28.5000	32.6000	1			
3	91	f	*	30.6000	1			
4	91	m	27.1000	34.3000	2			
5	91	m	27.8000	37.6000	3			
6	91	m	28.8000	39.0000	3			
7	91	f	28.0000	33.8000	2			
8								

8.6 Changing Data Types

Choose **Manip - Change Data Type**

This command is used to convert one data type into another. The possible conversions are as follows: Numeric to Text; Text to Numeric; Date/Time to Text; Date/Time to Numeric; Numeric to Date/Time; Text to Date/Time.

8.7 Sorting Data

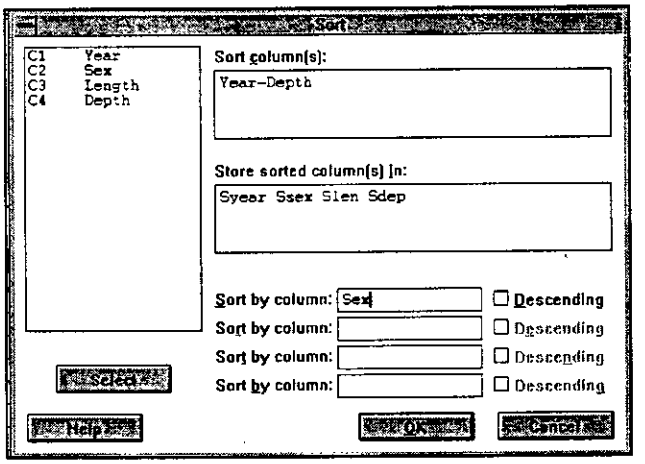
Choose **Manip - Sort**

This command is used to sort the data in a column. The default is in ascending order but values can also be sorted in descending order. Associated values of additional columns can also be carried along with the sorted column. Multiple sorts can also be done by first sorting a column depending on the values in one column, and then sorting within that sort by the values in a second column and so on.

For illustration we use the worksheet puffin7.mtw which contains the first seven lines of the puffin data. We sort the data so that records for females occur first, i.e. arrange the sex codes in alphabetical order and carry along the Lengths and Depths.

Fill in the dialog box as follows.

Select Year Sex, Depth & Length for **Sort column(s)** box.
Click on **Store sorted column(s) in** box.
Type in Syear Ssex Slen Sdep.
Click on the first **Sort by column(s)** box
Select Sex
(**Descending** box should be unchecked).
Click **OK**

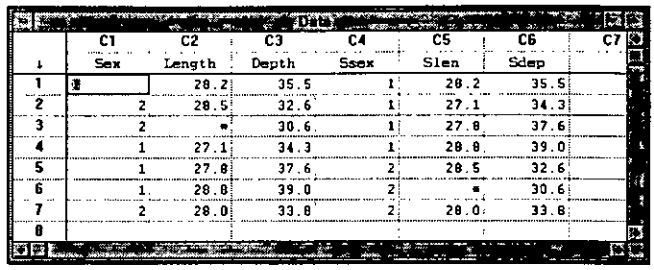


Session commands

```
MTB > Name C4 'Syear' C5 'Ssex' C6 'Slen' C7  'Sdep'
MTB > Sort  'Year'-'Depth' 'Syear' 'Ssex' 'Slen' 'Sdep';
SUBC> By Sex.
```

When sorting a single column the subcommand is not needed since the default is to sort on the first column.

The worksheet then contains

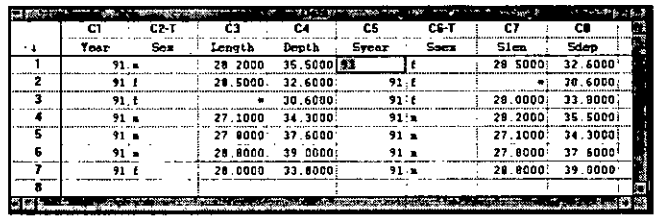


	C1	C2	C3	C4	C5	C6	C7
↓	Sex	Length	Depth	Ssex	Slen	Sdep	
1	1	28.2	35.5	1	28.2	35.5	
2	2	28.5	32.6	1	27.1	34.3	
3	2	*	30.6	1	27.8	37.6	
4	1	27.1	34.3	1	28.8	39.0	
5	1	27.8	37.6	2	28.5	32.6	
6	1	28.8	39.0	2	*	30.6	
7	2	28.0	33.8	2	28.0	33.8	
8							

Multiple sorting

This means sorting on the values of one column and then sorting values of another within the first sort. For example, in the puffin data after sorting on Sex we could sort the Lengths in increasing order. Starting from the above dialog box the steps are as follows. Click on the second **Sort by column** box. Select Length. Click **OK**.

The worksheet now shows that within each sex, lengths are arranged in ascending order.



	C1	C2-T	C3	C4	C5	C6-T	C7	C8
↓	Year	Sex	Length	Depth	Syear	Ssex	Slen	Sdep
1	91	m	28.2000	35.5000	93	f	28.5000	32.6000
2	91	f	28.5000	32.6000	91	f	*	30.6000
3	91	f	*	30.6000	91	f	28.0000	33.8000
4	91	m	27.1000	34.3000	91	m	28.2000	35.5000
5	91	m	27.8000	37.6000	91	m	27.1000	34.3000
6	91	m	28.8000	39.0000	91	m	27.8000	37.6000
7	91	f	28.0000	33.8000	91	m	28.8000	39.0000
8								

Missing data

Missing values (*) in numeric columns are sorted last (this is illustrated in the above example). Missing values in text data are sorted first.

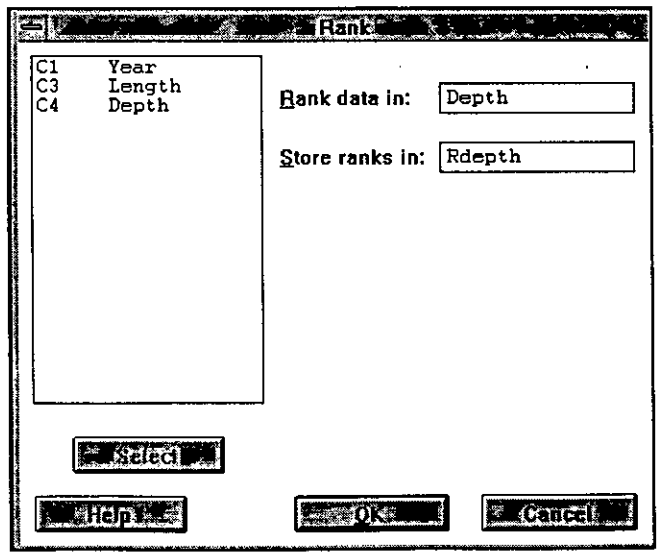
8.8 Ranking Data

Choose **Manip - Rank**

This assigns 1 to the smallest value, 2 to the next smallest and so on. Ties are assigned the average rank. The command RANK works with numeric columns only.

To rank the beak depths in the worksheet puffin7.mtw and store in Rdepth fill in the dialog box as follows.

Select Depth in **Rank data in** box.
Click on **Store ranks in** box.
Type in Rdepth.
Click **OK**



Session command

```
MTB > Name C4 'Rdepth'
MTB > Rank 'Depth'  'Rdepth'
```

The smallest depth is 30.6 mm, rank 1.
The second smallest is 32.6 mm, rank 2.

The largest is 39.0 mm, rank 7.

Data

	C1	C2-T	C3	C4	C5	C6
↓	Year	Sex	Length	Depth	Rdepth	
1	91	m	28.2000	35.5000	5	
2	91	f	28.5000	32.6000	2	
3	91	f	*	30.6000	1	
4	91	m	27.1000	34.3000	4	
5	91	m	27.8000	37.6000	6	
6	91	m	28.8000	39.0000	7	
7	91	f	28.0000	33.8000	3	
8						

Non-parametric methods

Ranks are required for some non-parametric methods, e.g. Spearman's rank correlation r_s. Rather surprisingly Minitab does not have a command for this. To calculate r_s for two variables such as length and depth, first rank them and then use the CORRELATE command to calculate a Pearson correlation on the ranks. If any rows of data contain missing values these must be removed before ranking the variables.

8.9 Displaying Data in the Session Window

Choose **Manip - Display Data**

This is used to display the value of specified columns, constants and matrices in the Session window. Note that before saving the results of a session it is usually a good idea to include a copy of the data used in the analysis.

Session command

For the puffin data

```
MTB > Print  'Year'   'Sex'  'Length'   'Depth'
```

9 Calculations

The **Calc** menu contains commands for calculations using both mathematical and logical expressions, functions, and for calculations on columns and across columns (i.e. rowwise). There are also commands for simulating random data, calculations with probability distributions, generating patterned data and matrix manipulations.

9.1 Mathematical Expressions and Functions

Choose **Calc - Calculator**

This is used for calculations of expressions containing arithmetic operations and mathematical functions, and for logical operations. It can also be used to access the value in a particular row of a column. The Minitab Session commands are assignment statements of the form

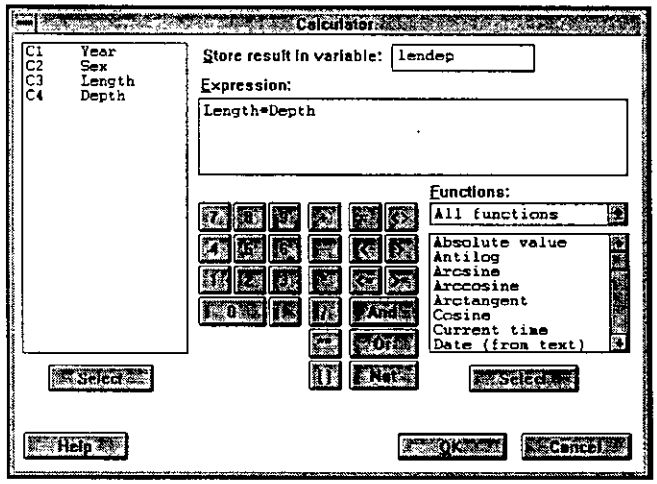
```
MTB > Let E = Expression
```

where E can be a column, a constant or a particular row of a column, and Expression may be algebraic or logical involving mathematical functions.

Example 1: to calculate an index of puffin beak size

To calculate an index of puffin beak size defined as the product of length and depth, fill in the dialog box as follows.

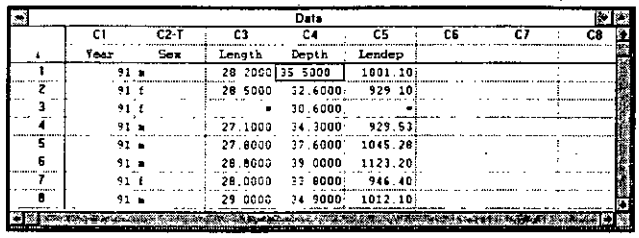
Click **Variable** box
Type in LenDep for the column name
Click **Expression** box
Type in the expression
Length * Depth (or C3*C4)
Click **OK**



Session command

```
MTB > Name C5 'Lendep'
MTB > Let 'Lendep' =  Length*Depth
```

The values of the calculated variable Lendep are stored in the worksheet as shown. For the row with a missing value for length the calculated index is also assigned a missing value.



Data

	C1	C2-T	C3	C4	C5	C6	C7	C8
	Year	Sex	Length	Depth	Lendep			
1	91	m	28.2000	35.5000	1001.10			
2	91	f	28.5000	32.6000	929.10			
3	91	f	*	30.6000	*			
4	91	m	27.1000	34.3000	929.53			
5	91	m	27.8000	37.6000	1045.28			
6	91	m	28.8000	39.0000	1123.20			
7	91	f	28.0000	33.8000	946.40			
8	91	m	29.0000	34.9000	1012.10			

Calculate can also be used for mathematical and statistical functions

Example 2: to calculate the natural log of beak length:

Type Ln_Len or C5 in **Store result in variable** box. Type loge(Length) or loge (C3) in the **Expression** box. Click **OK**

Or use the Session command:

```
MTB > Name C5 'ln_Len'
MTB > Let 'ln_Len' = loge(Length)
```
or
```
MTB > Let C5 = loge(C3)
```

Note that when a column name appears as an argument in a function it does not need to be enclosed in quotes.

Example 3: to calculate mean beak depth and store as a constant (K10)

Type K10 in **Store result in variable box.** Type mean(Depth) in the **Expression** boc. Click **OK.**

Or use session command:

```
MTB > Let k10 = Mean(Depth)
```

Comparison and Logical Operations (Subsetting Data)

These can be used to set up a column with row values to indicate whether some condition is true (row value set to 1) or false (row value set to 0). The columns used in the comparison must be numeric. For example, suppose we wanted to set up a column (say C5) to indicate males with beak depths greater than 35mm.

First, set up a numeric column for sex (Nsex say) with males coded as 1 and females as 2, using **Manip - Code - Text to Numeric** to code Sex as Nsex.

Then, enter C5 in the **Variable** box and in the **Expression** box, type 'Nsex' = 1 and 'Depth' > 35. Click **OK**

Or use the Session command:

```
MTB > Let C5 = 'Nsex' = 1 and 'Depth' > 35
```

The first few lines of the worksheet are then as shown.

Data

	C1	C2-T	C3	C4	C5	C6	C7	C8
	Year	Sex	Length	Depth	Nsex			
1	91	m	28.2000	35.5000	1			
2	91	f	28.5000	32.6000	2	0		
3	91	f	*	30.6000	2	0		
4	91	m	27.1000	34.3000	1	0		
5	91	m	27.8000	37.6000	1	1		
6	91	m	28.8000	39.0000	1	1		
7	91	f	28.0000	33.8000	2	0		

This could than be used with **Manip - Copy** toset up columns containing data for males with beak depths > 35.

Accessing Individual Elements of Columns

An individual element of a column can be accessed by enclosing the row number in brackets after the column number, e.g. C1(3) is the value in column 1 row 3. This is a single number (i.e. a Minitab constant). Individual elements can appear on both sides of a mathematical expression.

To store mean puffin beak depth in the first row of column C5

```
MTB > Let C5(1) = Mean(Depth)
```

The row number can be specified as a Minitab constant (e.g. K1), but the constant must have a value at the time. This is useful for writing Minitab macros.

9.2 Column Statistics

Choose **Calc - Column Statistics**

These commands calculate a statistic over the rows of a specified column. The result is printed out in the Session window but it can also be stored in a specified constant (e.g. K1). The available statistics are: Sum, Mean, Standard deviation, Minimum, Maximum, Range, Median, Sum of squares, Total N, Non-missing N and Missing N. Missing values are ignored in the calculations.

9.3 Row Statistics

Choose **Calc - Row Statistics**

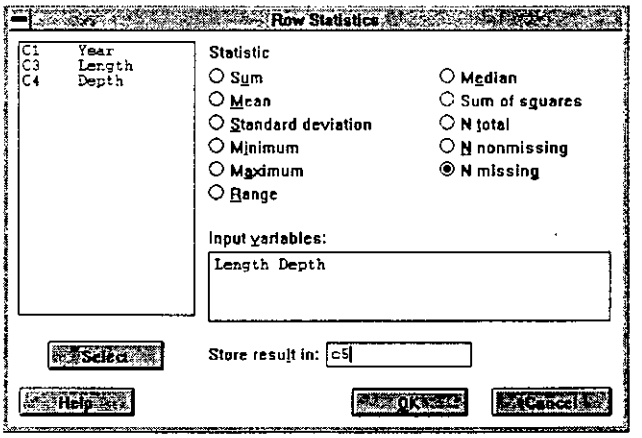
These commands calculate the value of a statistic across the rows of a specified set of columns (i.e. rowwise) The available list of statistics is the same as for **Column - Statistics**). The result is a column which is not printed but can be stored.

Indicating and removing rows with missing data

In some situations it could be required to remove rows of data containing one or more missing values. This can be done by setting up a column containing the number of missing values in the row and then using **Manip - Copy Columns - Use Rows** to select the complete rows.

Check on **N missing** box
Click **Input variables** box
Select 'Length' 'Depth'

Click **Store result in box**. Type in C5
Click **OK**



Session command

```
MTB > Rnmissing 'Length' 'Depth' C5
```

All commands for rowwise calculations begin with a letter R (upper or lower case).

The first few lines of the worksheet are as shown.

Data

↓	C1 Year	C2-T Sex	C3 Length	C4 Depth	C5	C6
1	91	m	28.2000	35.5000	0	
2	91	f	28.5000	32.6000	0	
3	91	f	*	30.6000	1	
4	91	m	27.1000	34.3000	0	
5	91	m	27.8000	37.6000	0	
6	91	m	28.8000	39.0000	0	
7	91	f	28.0000	33.8000	0	

9.4 Generating Patterned Data

Choose **Calc - Make Patterned Data**

This command allows data with some pattern or order to be entered in a particularly convenient way. There are options for sequences of Numeric data and Date/Time values.

Example 1: Randomised block layout

Suppose that in a randomised block experiment there are three treatments arranged in six blocks, and that the measured responses are entered into a particular column in the following order.

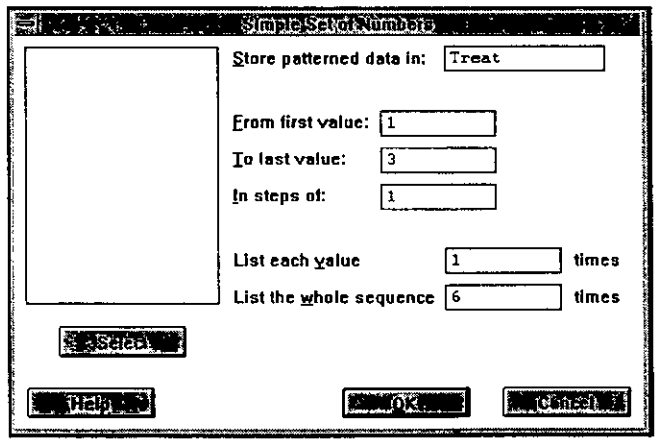
Treatment	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Block	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6

i.e. Block 1 first, followed by Block 2 and so on. Columns containing numerical codes for Treatments and Blocks can be set up by filling in the dialog box as follows.

Choose **Calc - Make Patterned Data - Simple Set of Numbers**

Enter a column by name or number in **Store patterned data in** box
Click **From first value** box.
Type in 1
Click **To last value** box.
Type in 3
Click **List the whole sequence** box.
Type in 6

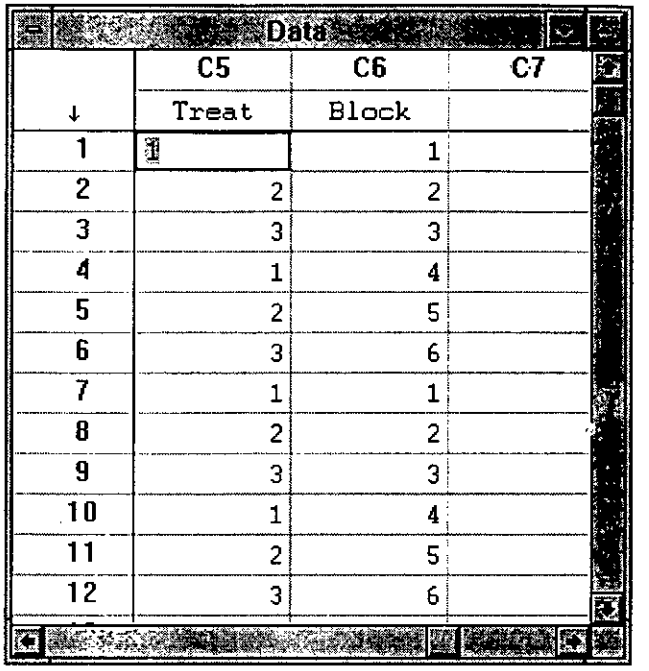
Click **OK**



A column containing codes for blocks can be set up in a similar way.

Session commands

```
MTB > Name C5 'Treat'
MTB > Set 'Treat'
DATA> 6(1:3)1
DATA> End
MTB > Name C6 'Block
MTB >  Set 'Block'
DATA> (1:6)3
DATA> End
```



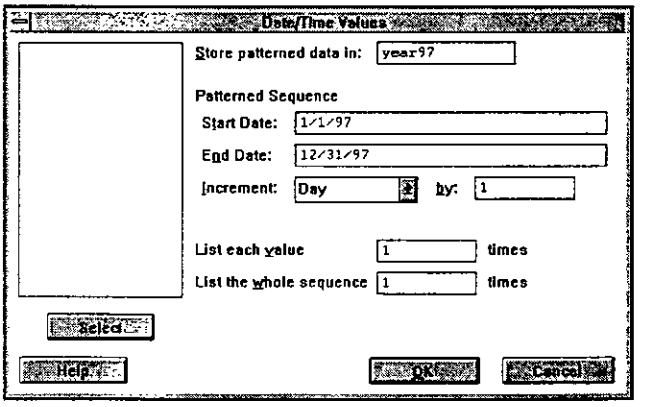
↓	C5 Treat	C6 Block	C7
1	1	1	
2	2	2	
3	3	3	
4	1	4	
5	2	5	
6	3	6	
7	1	1	
8	2	2	
9	3	3	
10	1	4	
11	2	5	
12	3	6	

Example 2: Generating date/time data

To create a column containing the days for 1997 in Date/Time format proceeds as follows.

Choose **Calc - Make Patterned Data -Date/Time Values**

Enter name of column (c5 or Year97) in **Store patterned data in** box
Check **Patterned sequence**
Enter 1/1/97 in **Start Date**
Enter 12/31/97 in **End Date**
(note that the month must be placed first)
The other default seetings apply
Click **OK**

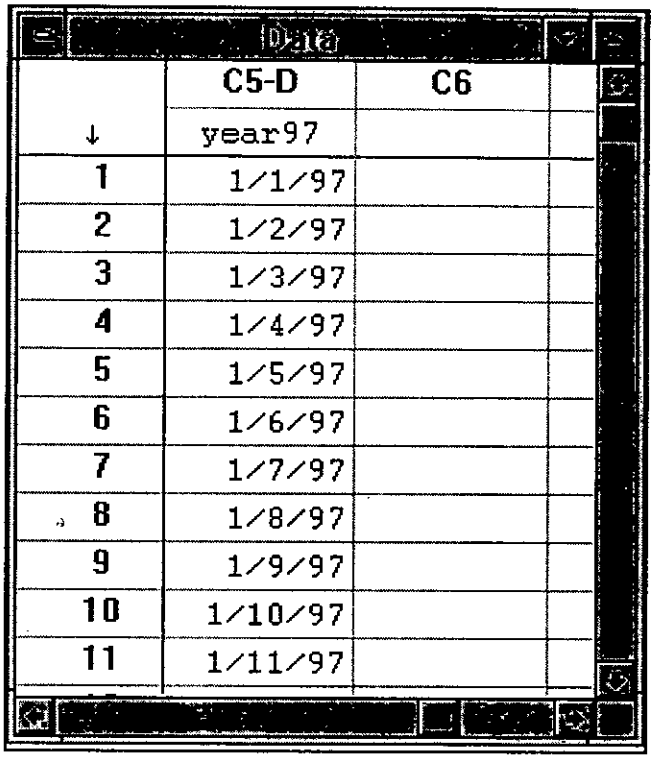


The date/time format can be changed using **Editor - Format Column - Date/Time.**

Session commands

```
MTB  > Dset 'Year97';
SUBC > Dstart '1/1/97';
SUBC > Dend '12/31/97';
SUBC > Day 1;
SUBC > Rvalue 1;
SUBC > Rlist 1.
```

The output is as shown.



Data

	C5-D	C6
↓	year97	
1	1/1/97	
2	1/2/97	
3	1/3/97	
4	1/4/97	
5	1/5/97	
6	1/6/97	
7	1/7/97	
8	1/8/97	
9	1/9/97	
10	1/10/97	
11	1/11/97	

9.6 Extracting from Date/Time Data

Choose **Calc - Extract Date/Time to Numeric** or **Calc -Extract Date/Time to Text**

This is used to extract the components of Date/Time data and store them in another column. For, example, suppose that a column contains dates of observations taken over one year in Month/Day/Year format. Then we can use **Extract Date/Time** to set up a column which gives the month of each observation either as numeric (1...12) or as text (Jan...Dec). This could then be used in tabulation or analysis of the observations by month.

9.7 Simulating Random Data

Choose **Calc - Random Data**

This is used to sample randomly from a column of data (with or without replacement) and to simulate random data from a range of distributions including the Normal, Binomial and Poisson. Possible applications are randomisation of treatments in a designed experiment, exploring the behaviour of statistical models and the sampling distributions of estimators, use in writing Minitab macros (see Section 13) for resampling methods such as randomisation tests and bootstrapping.

Example 1: Random selection of lottery numbers

For most statisticians the expected payoff from the lottery is not worth the gamble. However, for the more fun-loving types who like a flutter the following Minitab commands could be used to generate that lucky ticket (a 1 in 13,983,816 chance).

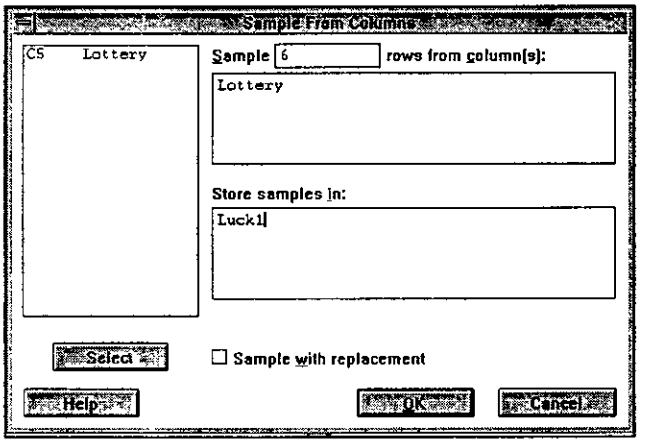
Set up a column containing the numbers 1-49 as follows.

```
MTB >   Set c5
DATA > 1:49
DATA > End
MTB >   Name c5 'Lottery'
```

Choose **Calc - Random Data - Sample from Column**

Fill in the dialog box as shown.

Note that in this example we are sampling without replacement (no duplicate numbers !) so that the **Sample with replacement** box is unchecked



Session command

```
MTB > Sample 6 c5 c6
```

The commands could be repeated 5 times to produce a batch of ticket numbers as shown.

Data

	C5	C6	C7	C8	C9	C10	C11
	Lottery	Luck1	Luck2	Luck3	Luck4	Luck5	
1	1	42	14	13	32	18	
2	2	25	45	15	23	34	
3	3	15	5	10	24	10	
4	4	48	46	23	5	6	
5	5	3	13	28	11	20	
6	6	16	31	18	41	44	
7	7						
8	8						
9	9						
10	10						

Example 2: Random shuffling

The approach in Example 1 can be used to randomly shuffle a set of numbers, as required in randomisation of treatments in a designed experiment. For example, in a randomised block design with 6 treatments, store the numbers 1-6 in c50 and use

```
MTB > Set c5
DATA> 1:6
DATA> End
MTB > Sample 6 c5 c6
```

This produces a random allocation for one block. Repeat the commands to obtain independent allocations for the other blocks.

Example 3: Simulating data from a Normal distribution

Choose **Calc - Random Data - Normal**

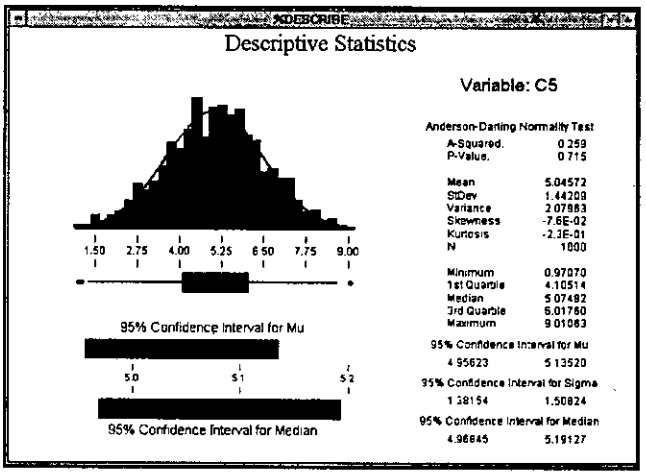
Fill in the dialog box specifying the number of observations, the column for storage, the mean and standard deviation of the distribution. For a sample of 1000 observations with mean 5 and standard deviation 1 stored in c5, the session commands are as follows.

```
MTB    > Random 1000 c5;
SUBC > Normal 5 1.5.
```

The graphical summary of the simulated data was obtained using

Stats - Basic Statistics -Descriptive Statistics
Graphs - Graphical Summary

This shows excellent agreement with a Normal distribution with mean 5 and standard deviation 1.5



Calculations on probability distributions

Choose **Calc - Probability Distributions**

This is used to calculate a probability density function (PDF), a cumulative distribution function (CDF) or the inverse (INVCDF) from a range of distributions. Calculations for the CDF are useful for calculating P-values associated with some statistical tests. Note however that in many analyses, for example regression and analysis of variance, Minitab provides P-values in the output.

For illustration, suppose that a chi-square test for differences in 3 proportions gives a value of 6.50. The P-value is the probability of observing a value of chi-square with 2 degrees of freedom as large or larger than 6.50. The cumulative distribution function (CDF) is the probability of a value less than or equal to 6.50, i.e 1 minus the required P-value. This is obtained as follows.

Choose **Calc - Probability Distributions - Chisquare**

Select Cumulative Probability. Enter Degrees of freedom (2) and Input constant (6.50). Specify a constant for the output (k10) - if omitted the output is displayed in the Session window. The P-value is 1 - k10.

Session commands

```
MTB  > Cdf 6.50 k10;
SUBC > Chisquare 2.
MTB  > Let k11 = 1 - k10
MTB  > Print k11
```

Conversely, for a given probability the corresponding lower percentage point can be obtained from the inverse of the cumulative distribution function (INVCDF). Fill out the dialog box or type in the following session commands.

```
MTB  > Invcdf  0.95;
SUBC > Chisquare 2.
```

The upper 5% point of the chisquare distribution on 2 degrees of freedom (5.992) will be displayed in the Session window.

10 Basic Statistics

The **Stat - Basic Statistics** menu contains commands for calculating summary statistics , t-tests, confidence intervals and the Pearson correlation coefficient.

10.1 Descriptive Statistics

Choose **Stat - Basic Statistics - Descriptive Statistics**

This produces a set of descriptive statistics for the data in a specified column, or for subsets of the data with separate statistics for each of the values held in some other specified column. Section 4 on the use of dialog boxes gives an example showing the calculation of descriptive statistics of the puffin beak measurement data separately for males and females.

The DESCRIBE command cannot be used to store the values of the calculated statistics - use STATS instead.

10.2 STATS

STATS is a Session command only. It computes a set of statistics on entire columns or subsets of columns (using the BY subcommand), and also has subcommand options for storing the results. It includes some statistics not included in the DESCRIBE command - for example skewness, kurtosis and the range. See HELP for the list of statistics calculated by STATS.

The following Minitab commands calculate the number of non-missing observations, the mean and standard of puffin beak length and depth. The results are stored in C6-C8 for length and C9-C11 for depth, with group codes in C5.

```
MTB >  Stats 'Length'  'Depth';
SUBC>  By  'Sex';
SUBC>  Gvalues C5;
SUBC>  N     C6 C9;
SUBC>  Mean  C7 C10;
SUBC>  Stdev C8 C11.
```

The results are stored in the worksheet as shown. No output is printed in the Session window.

Data

	C5-T	C6	C7	C8	C9	C10	C11	C12
1	M	21	28.6429	1.00079	21	35.3429	1.41724	
2	F	19	28.1789	0.93724	20	32.1650	1.68594	
3								
4								

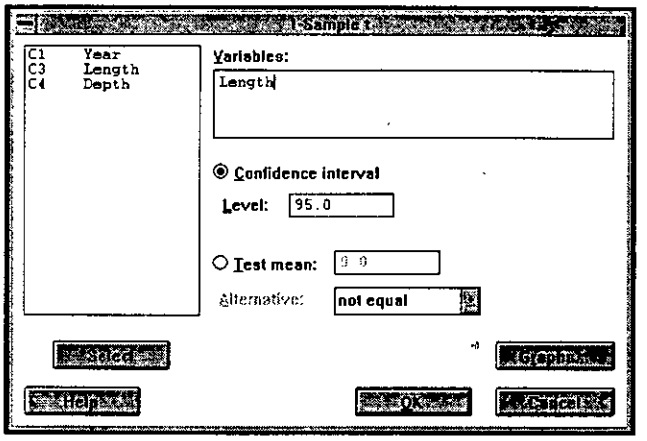
10.3 One-Sample t Confidence Interval for the Mean

Choose **Stat - Basic Statistics - 1-Sample t**

This can be used to calculate a 95% confidence interval for a population mean using a random sample from the population.

To obtain a 95% confidence interval for puffin mean beak length with puffin.mtw in the worksheet, complete the dialog box as follows.

Select Length
Click **OK**



Session command

```
MTB  > TInterval 95.0 'Length'.
```

Output

```
Confidence Intervals

Variable      N      Mean     StDev  SE Mean          95.0 % C.I.
Length       40    28.423     0.987     0.156  (  28.107,  28.738)
```

Interpretation

From a random sample of size N, the 95% confidence interval is calculated from the data using the formula: Mean +/- $t_{0.025,(N-1)}$ x (SE Mean), where $t_{0.025,(N-1)}$ is the upper 2.5% percentage point of Student's t-distribution with (N-1) degrees of freedom. The interval has the property that in repeated random samples from the same population the calculated confidence interval will contain the true population mean in *approximately* 95% of cases. We say *approximately* because the 95% confidence level is exact only when the population distribution is Normal. However, provided that the sample size is not too small the approximation is reasonable whatever the form of the population distribution.

In the above example we have ignored the fact that the sample contains a mixture of males and females (although it still could be a random sample). However, if there were marked differences in beak length between males and females it would be better to calculate a separate interval for each sex. This can be done by using **Manip - Copy** to copy beak lengths for each sex into two columns and then proceeding as above.

10.4 One-Sample t-test for the Mean

Choose **Stat - Basic Statistics - 1-Sample t**

The **1-Sample t** dialog box has an option to test the null hypothesis that the mean of a population distribution is some specified value.

For example, suppose that in a study of puffins on the Isle of May shows mean beak length to be 29mm - in a large sample of birds so that sampling error can be ignored. To test the null hypothesis that mean beak length of birds at St Kilda is equal to 29mm, fill in the dialog box as follows: Select Length. Click in **Test mean:** box. Type in 29. Click **OK**

The **Alternative** box is used to specify the alternative hypothesis - the default is **not equal** which means that there is no a priori specified direction for the difference, so that a two-tailed test is carried out.

Session commands

```
MTB > TTest 29.0 'Length';
SUBC>   Alternative 0.
```

Output

T-Test of the Mean

```
Test of mu = 29.000 vs mu not = 29.000

Variable      N      Mean     StDev    SE Mean         T     P-Value
Length       40    28.423     0.987      0.156    -3.70      0.0007
```

Interpretation

The P-value is the probability of observing a value of t as or more extreme than -3.70 ($-3.70 < t < 3.70$) when the null hypothesis is true. Since this probability is small we reject the null hypothesis and conclude that puffins on St. Kilda have shorter beaks than those on the Isle of May.

10.5 Two-Sample t-test to Compare Means

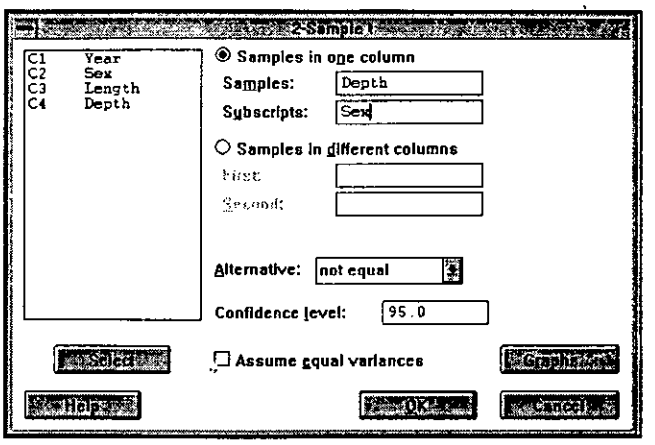
Choose **Stat - Basic Statistics - 2-Sample t**

This tests for differences between two population means using random samples from each population. It also produces a confidence interval for the difference between population means. Data can be presented either as a single column containing the responses with a corresponding

column identifying sample origins, or with the responses for the two samples in different columns.

To test for differences between mean beak measurements of male and female puffins and to calculate a 95% confidence interval fill in the dialog box as follows (puffin.mtw in worksheet).

Check **Samples in one column**
Click in **Samples** box and select Depth
Click in **Subscripts** and Select Sex
Set **Alternative** as 'not equal to'
Set **Confidence level** to 95%
Click **OK**



When **Assume equal variances** is checked the method assumes that the population variances are equal, and a pooled estimate is used to estimate the standard error of the difference between the sample means. The default is not to assume equal variances so that separate standard errors are calculated for each sample mean. This method is more robust when the population variances differ.

Session commands

```
MTB > TwoT 95.0 'Depth' 'Sex';
SUBC>   Alternative 0.
```

Output

Two Sample T-Test and Confidence Interval

```
Twosample T for Depth
Sex   N      Mean      StDev   SE Mean
m    21     35.34       1.42       0.31
f    20     32.17       1.69       0.38

95% C.I. for mu 1 - mu 2: ( 2.19,  4.17)
T-Test mu 1 = mu 2 (vs not =): T= 6.52  P=0.0000  DF=  37
```

Interpretation

The bottom line of the output shows a small P-value and strong evidence against the null hypothesis of no difference between males and females. This 95% confidence interval for the

actual difference is 2.19-4.17mm (see previous section for interpretation of the P-value and 95% confidence interval).

10.6 Confidence Interval and t-test for Paired Data

In some situations the data occur in pairs - for example, observations on a sample of individuals before a treatment is applied and after the treatment has been applied. In this case we work with the difference and use the 1-Sample methods described above.

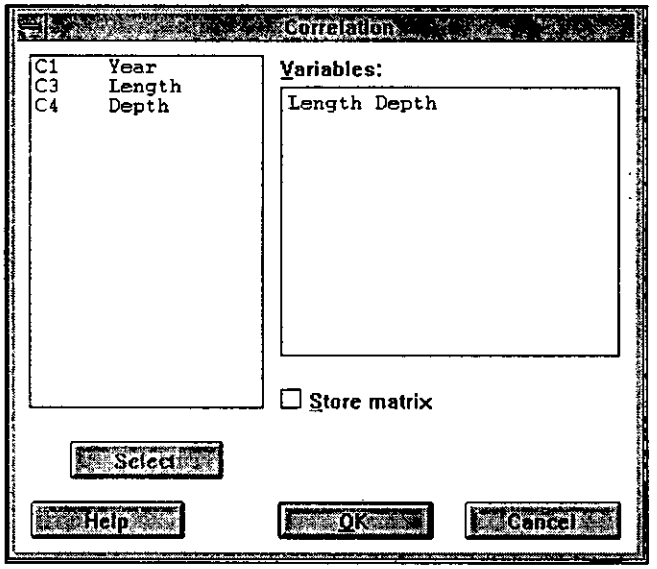
10.7 Correlation

Choose **Stat - Basic Statistics - Correlation**

To calculate the correlation between beak length and beak depth fill in the dialog box as follows.

Select Length and Depth
Click **OK**

If the **Store matrix** box is checked the correlation coefficient(s) will be stored as a matrix



Session command

```
MTB> Correlation 'Length' 'Depth'
```

Output

```
Correlations (Pearson)

Correlation of Length and Depth = 0.162
```

Interpretation

Minitab calculates the correlation coefficient excluding pairs with any missing values from either variable, but the output does not show the number of pairs - in this case, there are 41 pairs of Length and Depth values but since one value of Length is missing the coefficient is calculated from 40 pairs.

The correlation coefficient (r) is used to test for a relationship between two variables, i.e. the null hypothesis is zero correlation, but the P-value for this test is not provided by Minitab. This can be obtained from tables or by calculating $t = r*\text{sqrt}[(n-2)/(1 - r^2)]$ where n is the number of non-missing pairs of observations, and referring t to Student's t-distribution with (n-2) degrees of freedom. In this case, $t = 1.01$ ($P = 0.32$).

***** Important *****

With more than two variables Minitab calculates a matrix of pairwise coefficients, using all available data for each pair of variables. Sample sizes for the coefficients may therefore vary depending on the pattern of missing observations. This contrasts with some other statistical packages which omit whole rows of data with one or more missing value before calculating the correlation matrix.

****

Spearman's rank correlation coefficient

Spearman's rank correlation coefficient is calculated by ranking both columns of data and then using CORRELATION on the ranked data. Any rows which contain missing data should be removed first, otherwise the ranks in the two columns will not be comparable (see above section on Row Statistics). Session commands which could be used to calculate Spearman's rank correlation between beak length and depth are as follows.

```
MTB > Rnmiss('Length' - 'Depth') c5
MTB > Copy 'Length' 'Depth' C6 C7;
SUBC>  Use C5 = 0.
MTB > Rank C6 C8
MTB > Rank C7 C9
MTB > Correlation C8 C9
```

11 Tables

Choose **Stat - Tables**

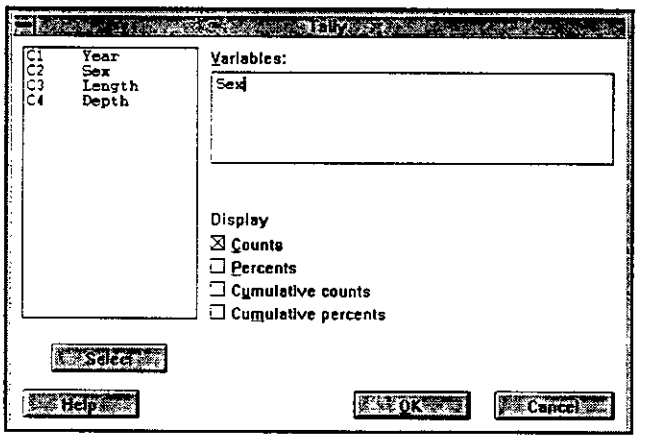
The TABLES command is used to tabulate one-way, two-way and multi-way tables of summary statistics. It can also be used to calculate chi-square tests for contingency tables. There are options on the layout for printing the tables.

11.1 Tally

Choose **Stat - Tables - Tally**

Tally prints a summary table of the frequency of values in one or more columns. The columns can be Text or Numeric (containing integers from -9999 to +9999), or the missing value code. Frequencies can be counts or percentages, or cumulative. To count the number of males and females in the puffin worksheet puffin.mtw:

Select Sex for **Variables** box.
Click **OK**



Session command

```
MTB > Tally 'Sex';
SUBC > Counts.
```

Output

```
Summary Statistics for Discrete Variables

     Sex  Count
       m     21
       f     20
     N=      41
```

For the Session command the tally frequencies can be saved in a column using the subcommand STORE.

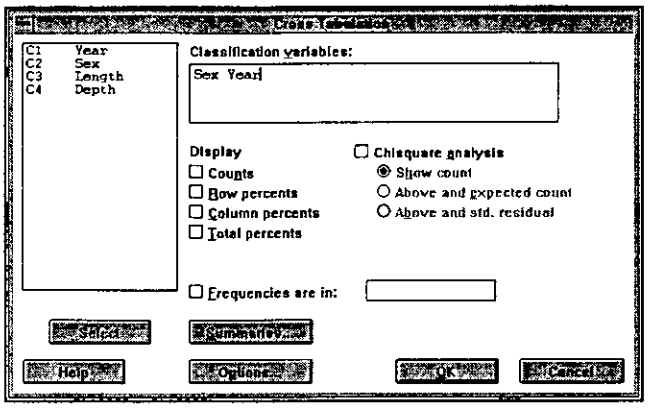
11.2 Cross-tabulation

Choose **Stat - Tables - Cross Tabulation**

This command calculates one-way, two-way and multi-way tables of summary statistics.

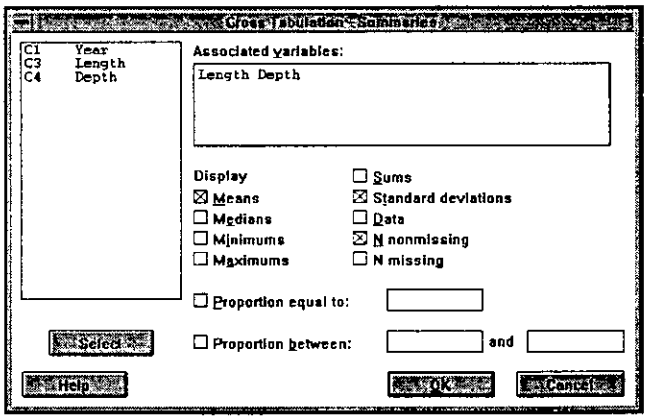
To tabulate by year and sex the mean, standard deviation and number of non-missing values for Length and Depth fill in the dialog boxes as follows.

Select Sex Year as **Classification variables.**
Click **Summaries** box to go to
Cross Tabulation Summaries box



Select Length Depth for
Associated variables box.
Check **Display** boxes for:
Means, Standard deviations,

N nonmissing.
Click **OK**
(returns to **Cross Tabulation** box).
Click **OK.**



Session commands

```
MTB > Table 'Sex' 'Year';
SUBC>   Means 'Length' 'Depth';
SUBC>   StDev 'Length' 'Depth';
SUBC>   N 'Length' 'Depth'.
```

Output

Tabulated Statistics

ROWS: Sex COLUMNS: Year

	91	92	93	ALL
m	28.150	28.400	29.225	28.643
	36.367	34.986	34.888	35.343
	0.692	1.105	0.905	1.001
	1.784	1.067	1.108	1.417
	6	7	8	21
	6	7	8	21
f	28.000	28.100	28.280	28.179
	31.525	33.067	31.880	32.165
	0.500	0.858	1.125	0.937
	2.087	1.172	1.724	1.686
	3	6	10	19
	4	6	10	20
ALL	28.100	28.262	28.700	28.423
	34.430	34.100	33.217	33.793
	0.606	0.970	1.114	0.987
	3.078	1.461	2.108	2.223
	9	13	18	40
	10	13	18	41

CELL CONTENTS --
Length:MEAN
Depth:MEAN
Length:STD DEV
Depth:STD DEV
Length:N
Depth:N

Controlling the layout of the printed table

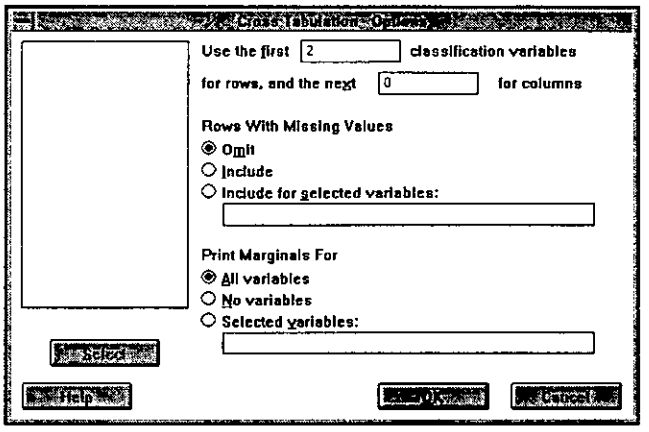
In the above output the summary statistics for the different variables are stacked in the cells of the table. The layout can be changed so that summary statistics for different variables are printed side by side. This is done by using the **Options** box from the **Cross Tabulation** menu as follows.

To augment the above example:

Click on **Options** in the **Cross Tabulation** box.
Click on the first box. Type in 2
Click on the second box. Type in 0

Click **OK** to return to **Cross Tabulation** box.

Click **OK**



Session commands

```
MTB > Table 'Sex' 'Year';
SUB >  Layout 2 0;
SUBC>   Means 'Length' 'Depth';
SUBC>   StDev 'Length' 'Depth';
SUBC>   N 'Length' 'Depth'.
```

Output

Tabulated Statistics

ROWS: Sex / Year

		Length MEAN	Depth MEAN	Length STD DEV	Depth STD DEV	Length N	Depth N
m							
	91	28.150	36.367	0.692	1.784	6	6
	92	28.400	34.986	1.105	1.067	7	7
	93	29.225	34.888	0.905	1.108	8	8
f							
	91	28.000	31.525	0.500	2.087	3	4
	92	28.100	33.067	0.858	1.172	6	6
	93	28.280	31.880	1.125	1.724	10	10
ALL							
	ALL	28.423	33.793	0.987	2.223	40	41

12 More Advanced Topics

This Introduction to Minitab is largely concerned with using Minitab through menus and dialog boxes, input and output, manipulating and summarizing data and performing some simple statistics for comparing populations and testing for trends. However, Minitab has many more commands for carrying out a wide range of statistical analyses. A brief summary of some more advanced topics is given below.

12.1 Regression

Choose **Stat - Regression**

Minitab has excellent facilities for linear regression. The menu includes the following.

- Single variable and multiple regression models.
- Model selection procedures - forward, backward and stepwise methods.
- Best subsets regression.
- Weighted regression.

The output contains a comprehensive summary of the fitted model and a full set of regression diagnostics which can be saved to check the fit of the model.

New in Release 10

- X-Y scatterplot with fitted regression line, confidence bands and prediction bands.
- Residual diagnostic plot - including histogram and Normal plots of residuals, residuals versus fitted values, plots for detecting order effects and outliers.

New in Release 11

- Polynomial regression

- Logistic regression for analysing data where the response variables are binary (0/1) or proportions derived from binary scores (e.g. proportion of individuals surviving)

12.2 Analysis of Variance

Choose **Stat - ANOVA**

The menu includes the following.

- One-way ANOVA - with Tukey, Fisher, Dunnet and MCB multiple comparison procedures
- Two-way ANOVA - balanced designs.
- Mixed model ANOVA - for balanced designs with fixed and random effects, crossed and nested factors.
- Analysis of Covariance (ANCOVA) - for balanced designs with additive covariates.
- General Linear Models (GLM) - for unbalanced analysis of variance. This is a very flexible method. It should be emphasised that GLMs is not to be confused with Generalized Linear Models which can be used to analyse some types of data which can be markedly non-Normal - for example, count data for which a Poisson or Binomial distribution could be a plausible working model. GLMs are for models in which the response follows a Normal distribution, or at least approximately so.

As in the Regression commands the output from ANOVA contains a comprehensive summary of the fitted model and a set of ANOVA diagnostics which can be saved for checking the fit of the model

New in Release 10

- Residual diagnostic plots.
- Main effects plot - up to 50 factors.
- Interaction effects plot - two-way interactions up to 9 factors

12.3 Nonparametric Statistics

Choose **Stat - Non-Parametrics**

This menu includes a range of non-parametric statistics. These methods make no specific assumptions about the form of the distribution of the data. They use ranks, relative values and or signs of differences in place of the observations.

- 1-Sample sign test and confidence interval
- 1-Sample Wilcoxon signed rank test and confidence interval
- Mann-Whitney-Wilcoxon rank sum two-sample test to compare two populations, and confidence interval
- Kruskal-Wallis k-sample test to compare k populations
- Mood's median test to compare k populations
- Friedman test for randomised blocks
- Calculation of pairwise averages, differences and slopes.

Releases 10 and 11 don't contain any new nonparametric methods.

12.4 Multivariate Analysis

Choose **Stat - Multivariate**

The menu includes the following.

- Principal Components Analysis (PCA) - using either the correlation matrix or the variance-covariance matrix.
- Linear and quadratic discrimination analysis - including cross-validation to estimate the misclassification rate and the allocation of new individuals.
- Factor analysis

New in Release 10

- Cluster Analysis - including hierarchical and k-means analysis, dendrogram output.

New in Release 11

- Correspondence analysis–

13 Macros

Minitab is usually used interactively, i.e. commands are carried out as they are entered. However, Minitab commands can also be stored in a file and then executed as a macro. This can be useful for repeated calculations, simulations and more generally it provides a programming language. Minitab 11 macros are either Exec files or %Macro files. %Macros were introduced in Release 9, and they are more powerful and flexible than Execs which were available on earlier releases. A full discussion of %Macros is beyond the scope of the course.

13.1 Execs

Execs are files containing Minitab commands which can be executed in a Minitab session. They can be used for repeating blocks of calculations, looping through the columns and rows of the worksheet to perform analyses for different columns and subsets of the data, writing tailor-made programs. Execs can prompt the user for information during execution.

Example

The worksheet guill.mtw contains measurements on wing length (mm) and weight (g) in samples of guillemot chicks obtained in a study on the Isle of May during 1982-86 (coded as 82-86 in the column Year). Suppose that for each year we wanted to plot weight against wing length - the relationship is of interest because wing length could be used as a covariate in comparing chick weights in different years. The analysis could be done manually by successively copying weight (Wlength) and length (Length) for each of the five years into two columns and then plotting them out. Alternatively, it can be done using an Exec file by placing the following commands in a file plotguil.mtb as follows.

```
Copy 'Wlength' 'Weight' c51 c52;
 Use 'Year' = k50.
Name c52 'Wt' c51 'Wlen'
Plot c52*c51
Let k50 = k50+1
```

Columns are named for a more informative graph.

The constant k50 is a counter which successively takes the values 82,83,...,86 when the Exec is run. Initialise k50 and run the Exec as follows.

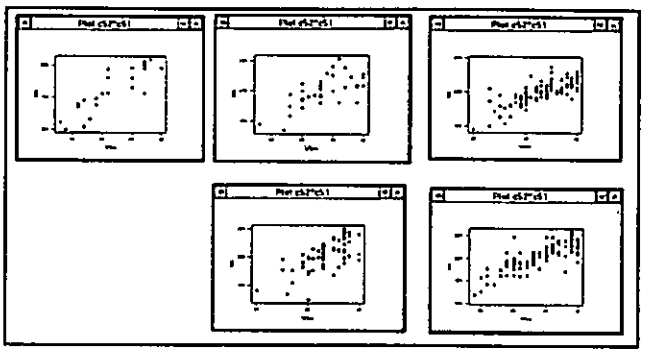
```
MTB > Let k50 = 82
MTB > Exec 'c:\intmtb11\plotguil'  5
```

Or Choose **File - Other Files - Run an Exec**

Specify the number of times to run the Exec (in this case 5) and then select the Exec file name.

The resulting five graphs are shown sized and positioned in the order

1982 1983 1984
1985 1986



Creating Exec Files

This can be done in three ways.

1. Use an editor to create the file (with extension MTB)

2. In a Minitab session store the commands by using the JOURNAL command (section 7.6). The file will have an extension MTJ which needs to be specified when running the Exec.

3. Use the store command in a Minitab session as follows.

```
MTB   > Store 'c:\intmtb11\plotguil'
STORE > Copy 'Wtlength' ' Weight' c51 c52
STORE > Name c52 'Wt' c51 'Wlen'
STORE > Plot c52*c51
STORE > Let k50 = k50+1
STORE > End
```

Command Line Editor

Choose **Edit - Command Line Editor**

Commands can be entered directly into the Command Line Editor or copied from the Session or History window. They can then be edited before submitting. This provides an easy way of developing the code for a Minitab macro.

Looping Through Columns

Minitab has a so-called CK capability which allows looping through columns. For example, suppose that we have 6 columns (c1-c6) and we want to plot c6 against c1-c5 in turn. This can be done using an Exec file plot5.mtb containing the following commands.

```
Plot c6*ck50
Let k50 = k50+1
```

Initialise k50 and run the Exec as follows.

```
MTB > Let k50 = 1
MTB > Exec 'plot5'  5
```

13.2 Global and Local Macros (%Macros)

%Macros were introduced in Release 9. They can be either global or local but both types have default file extension MAC. Global macros are generally simpler and easier to write than local macros, but are less powerful and flexible. For example, in local macros columns and options can appear as arguments, and variables may be local to the macro so avoiding overwriting and cluttering up the worksheet.

The general structure of a global macro is as follows.

```
Gmacro
Template (i.e. name of macro)
Body of macro (i.e. Minitab commands)
Endmacro
```

An example which performs the same function of the Exec 'plotguil.mtb' discussed in the previous section is the following global macro plotguil.

```
Gmacro
Plotguil
Name c51 'Wlen' c52 'Wt'
Do k50=82:86
 Copy 'Wlength' 'Weight' c51 c52;
   Use   'Year' = k50.
 Plot c52*c51
Enddo
Endmacro
```

To execute the macro from a Minitab session, type %plotguil after the Minitab prompt.

In a local macro the columns for Wlength, Weight and Year could be passed from the worksheet as arguments.

A full discussion of %Macros is beyond the scope of this course. For more details see the Minitab Reference Manual.

14 Documentation

MINITAB User's Guide, Release 11 for Windows. A useful introduction with step-by-step sample sessions.

MINITAB Reference Manual, Release 11 for Windows.

MINITAB Quick Reference, Release 11 for Windows. A summary of Minitab menus, dialog boxes, windows and commands.

15 Minitab 11 - some comments on differences

For users of Minitab Release 10, the following new features of Minitab Release 11 are worth noting.

- **Session window default change.** The session window displays only text output generated by analysis commands. The default setting can be changed using Edit - Save Preferences.

- **Context-sensitive shortcut menus.** Click the right mouse button to display context-sensitive 'fast' menus for convenient access to the most common operations.

- **Worksheet description.** Minitab 11 allows you to save some comments to be associated with a worksheet.

- **Insert/Delete columns from Data window.** It is now much easier to insert/delete columns from the data window.

- **Long column names.** Column names can be up to 31 characters long.

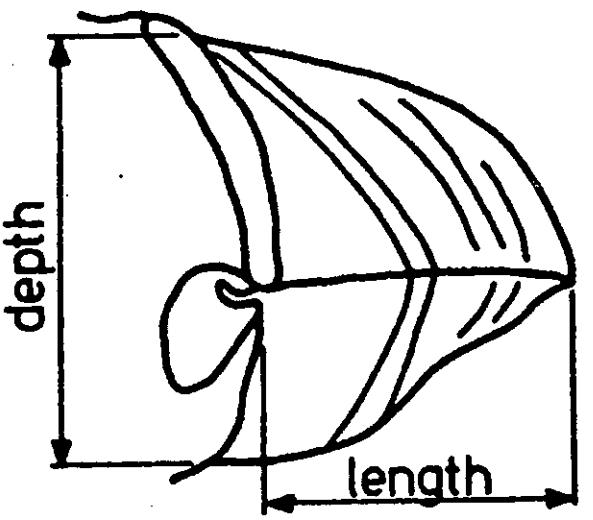
- **Alpha data.** In Minitab 11 most commands allow alpha columns (referred to as text columns) for specifying levels of some factor.

- **Date/Time data.** Minitab 11 can handle date/time variables.

- **Statistical methods.** Additional statistical methods include logistic regression, polynomial regression and correspondence analysis.

Appendix

```
#     Puffin beak measurements
#
#     c1 - Year
#     c2 - Sex
#     c3 - Length(mm)
#     c4 - Depth(mm)
#
91   m   28.2   35.5
91   f   28.5   32.6
91   f      *   30.6
91   m   27.1   34.3
91   m   27.8   37.6
91   m   28.8   39.0
91   f   28.0   33.8
91   m   29.0   34.9
91   f   27.5   29.1
91   m   28.0   36.9
92   m   29.7   35.9
92   m   26.6   34.4
92   f   28.0   34.0
92   m   28.2   34.6
92   f   29.0   34.7
92   m   27.7   34.0
92   m   28.0   37.0
92   f   27.6   33.2
92   f   29.3   31.4
92   f   27.5   32.5
92   f   27.2   32.6
92   m   29.1   34.5
92   m   29.5   34.5
93   f   30.4   32.8
93   f   29.1   34.0
93   f   27.8   33.8
93   f   27.6   32.0
93   m   30.0   36.1
93   m   27.6   36.5
93   f   26.6   31.9
93   m   28.8   34.3
93   f   28.0   28.6
93   m   30.5   34.1
93   f   29.2   29.8
93   m   28.8   33.5
93   m   29.9   35.8
93   f   28.6   32.3
93   m   29.2   34.8
93   f   27.0   32.9
93   f   28.5   30.7
93   m   29.0   34.0
```



ITE has six Research Stations throughout Britain, which allows the efficient use of resources for regional studies and provides an understanding of local ecological and land use characteristics. The Institute's administrative headquarters is at Monks Wood.

ITE sites

Monks Wood
(Admin HQ)
Abbots Ripton
HUNTINGDON PE17 2LS
Telephone 01487 773381-8
Fax 01487 773467
Email MONKSWOOD@ITE.AC.UK

Merlewood Research Station
GRANGE-OVER-SANDS
Cumbria LA11 6JU
Telephone 015395 32264
Fax 015395 34705
Email MERLEWOOD@ITE.AC.UK

Edinburgh Research Station
Bush Estate
PENICUIK
Midlothian EH26 0QB
Telephone 0131 445 4343
Fax 0131 445 3943
Email BUSH@ITE.AC.UK

Furzebrook Research Station
WAREHAM
Dorset BH20 5AS
Telephone 01929 551518-9, 551491
Fax 01929 551087
Email FURZEBROOK@ITE.AC.UK

Banchory Research Station
Hill of Brathens
Glassel, BANCHORY
Kincardineshire AB31 4BY
Telephone 01330 823434
Fax 01330 823303
Email BANCHORY@ITE.AC.UK

Bangor Research Unit
University of Wales, Bangor
Deiniol Road
BANGOR, Gwynedd LL57 2UP
Telephone 01248 370045
Fax 01248 355365
Email BANGOR@ITE.AC.UK

Details about the Institute are available on the Internet via the World Wide Web (http:/www.nmw.ac.uk/ite)